

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 15, 2004, 19:19:09 ; Search time 33 Seconds
(without alignments)
3971.533 Million cell updates/sec

Title: US-09-522-753-5
Perfect score: 13215
Sequence: 1 MSGSTQLVAQWTRATEPRYP.....WDEBPKPLCSQYTLSDSE 2517

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	13215	100.0	2517	1 NCR2_HUMAN	O9y618 h nuclear r
2	10987.5	83.1	2472	1 NCR2_MOUSE	Q9wu42 mus musculus
3	4187	31.7	2453	1 NCR1_MOUSE	Q60974 mus musculus
4	4147.5	31.4	2440	1 NCR1_HUMAN	O75376 homo sapien
5	714.5	5.4	533	1 NCR1_RAT	Q9wub5 rattus norv
6	576	4.4	3644	1 MINT_MOUSE	Q62504 mus musculus
7	569.5	4.3	3664	1 MINT_HUMAN	Q9st58 homo sapien
8	526	4.0	2142	1 BAT2_HUMAN	P48634 homo sapien
9	503	3.8	2716	1 OSA_DROME	Q8in94 drosophila
10	492.5	3.7	2774	1 MAPA_RAT	P34926 rattus norv
11	470.5	3.6	5262	1 MLL2_HUMAN	O14686 homo sapien
12	466.5	3.5	5560	1 SPEN_DROME	O8ax83 drosophila
13	462.5	3.5	5085	1 PCLO_RAT	Q9jks6 rattus norv
14	451.5	3.4	2805	1 MAPA_HUMAN	P78559 homo sapien
15	439.5	3.3	2067	1 NCO6_MOUSE	Q9j119 m nuclear r
16	439	3.3	4911	1 MLL3_HUMAN	Q8he24 homo sapien
17	438.5	3.3	1780	1 YKZ6_CABEL	P34333 caenorhabdi
18	434.5	3.3	5038	1 PCLO_MOUSE	Q9qyx7 mus musculus
19	432	3.3	2167	1 SHK1_RAT	Q9wv48 rattus norv
20	428.5	3.2	2715	1 MLL4_HUMAN	Q9umh6 homo sapien
21	428	3.2	2063	1 NCO6_HUMAN	O14686 h nuclear r
22	424.5	3.2	2464	1 MAPB_MOUSE	P14873 mus musculus
23	415.5	3.1	1226	1 SNT1_YEAST	P25357 saccharomyc
24	410.5	3.1	5147	1 PCLO_HUMAN	Q9y6v0 homo sapien
25	410	3.1	2468	1 MAPB_HUMAN	P46821 homo sapien
26	407	3.1	5120	1 PCLO_CHICK	Q9pu36 gallus gall
27	406.5	3.1	2459	1 MAPB_RAT	P15205 rattus norv
28	400.5	3.0	1464	1 CA11_HUMAN	P02452 homo sapien
29	399	3.0	2161	1 SHK1_HUMAN	Q9y566 homo sapien
30	399	3.0	4903	1 MLL3_MOUSE	Q8brh4 mus musculus
31	396	3.0	3703	1 ABP1_HUMAN	Q15911 homo sapien
32	391.5	3.0	1411	1 TCOF_HUMAN	Q13428 homo sapien
33	391.5	3.0	1460	1 CA11_CANPA	Q9xsj7 canis famil

ALIGNMENTS

RESULT 1

34	391	3.0	3924	1	ANK2_HUMAN	Q01484 homo sapien
35	390	3.0	2944	1	CA17_HUMAN	Q02388 homo sapien
36	389.5	2.9	3866	1	HRX_MOUSE	P55200 mus musculus
37	389	2.9	1670	1	CA34_HUMAN	Q01955 homo sapien
38	382.5	2.9	1362	1	BRD4_HUMAN	O60885 homo sapien
39	380	2.9	3969	1	HRX_HUMAN	Q03164 homo sapien
40	379	2.9	1685	1	CA54_HUMAN	P29400 homo sapien
41	378	2.9	1183	1	DRPL_RAT	P54258 rattus norv
42	375.5	2.8	3726	1	ABP1_MOUSE	Q61329 mus musculus
43	373	2.8	1763	1	CA24_ASCSU	P27393 ascaris suu
44	368	2.8	1185	1	DRPL_HUMAN	P54259 homo sapien
45	364	2.8	1426	1	BCL9_HUMAN	O00512 homo sapien
46	360	2.7	1509	1	BCL9_HUMAN	Q9nzm4 homo sapien
47	358	2.7	1618	1	NEST_HUMAN	P48681 homo sapien
48	355.5	2.7	2842	1	APC_RAT	P70478 rattus norv
49	355	2.7	2090	1	N214_HUMAN	P35658 homo sapien
50	353	2.7	1453	1	CA11_CHICK	P02457 gallus gall
51	353	2.7	1983	1	TF20_MOUSE	Q9epq8 mus musculus
52	352	2.7	1838	1	CA15_HUMAN	P20908 homo sapien
53	350.5	2.7	1736	1	CA2B_HUMAN	P13942 homo sapien
54	350.5	2.7	2843	1	APC_HUMAN	P25054 homo sapien
55	349.5	2.6	1822	1	ZAP5_HUMAN	P49750 homo sapien
56	348.5	2.6	1464	1	CA13_MOUSE	P08121 mus musculus
57	348	2.6	1669	1	CA14_MOUSE	P02463 mus musculus
58	348	2.6	3256	1	KI67_HUMAN	P46013 homo sapien
59	346	2.6	1338	1	ACIN_MOUSE	Q9jix8 mus musculus
60	344	2.6	2845	1	APC_MOUSE	Q61315 mus musculus
61	341.5	2.6	1026	1	NFH_HUMAN	P12036 homo sapien
62	341	2.6	1453	1	CA11_MOUSE	P11087 mus musculus
63	341	2.6	1815	1	SHK3_RAT	Q9jlu4 rattus norv
64	340	2.6	1466	1	CA13_HUMAN	P02461 homo sapien
65	339.5	2.6	1341	1	ACIN_HUMAN	Q9ukv3 homo sapien
66	339.5	2.6	1690	1	CA44_HUMAN	P53420 homo sapien
67	339	2.6	1960	1	TF20_HUMAN	Q9ugu0 homo sapien
68	338.5	2.6	1505	1	CUT2_HUMAN	O14529 homo sapien
69	337.5	2.6	2426	1	SON_HUMAN	P18583 homo sapien
70	335.5	2.5	5703	1	MUSB_HUMAN	Q9hc84 homo sapien
71	335	2.5	1516	1	CA1H_HUMAN	Q3060 homo sapien
72	334	2.5	1739	1	DOTL_HUMAN	Q8tek3 homo sapien
73	332.5	2.5	1902	1	SMF1_HUMAN	O14497 homo sapien
74	332.5	2.5	2404	1	SON_MOUSE	Q9qx47 mus musculus
75	332	2.5	810	1	NFM_BOVIN	O77788 bos taurus
76	330.5	2.5	1418	1	CA12_HUMAN	P02458 homo sapien
77	330.5	2.5	2349	1	TPR_HUMAN	P12270 homo sapien
78	330	2.5	1367	1	AMYH_YEAST	P08640 saccharomyc
79	330	2.5	1461	1	IF18_PRTIF	P11675 pseudorabie
80	329.5	2.5	915	1	NFM_HUMAN	P07197 homo sapien
81	329.5	2.5	1083	1	TFD3_HUMAN	O00268 homo sapien
82	325	2.5	1003	1	MBD6_HUMAN	Q96dn6 homo sapien
83	325	2.5	1781	1	AK12_HUMAN	Q02952 homo sapien
84	324	2.5	3149	1	TEGU_EBV	P03186 epstein-bar
85	323.5	2.4	1603	1	CA1F_HUMAN	Q07092 homo sapien
86	323	2.4	1446	1	IF18_PRTVA	P33479 pseudorabie
87	322.5	2.4	2004	1	MYS3_HUMAN	Q92794 homo sapien
88	320.5	2.4	1669	1	CA14_HUMAN	P02462 homo sapien
89	319.5	2.4	1362	1	CA21_CHICK	P02467 gallus gall
90	319	2.4	1736	1	CA2B_MOUSE	Q64739 mus musculus
91	317	2.4	555	1	GP1_CHLRE	Q9fpg6 chlamydomon
92	316	2.4	1603	1	PSC_DROME	P35820 drosophila
93	315.5	2.4	1459	1	CA12_MOUSE	P28481 mus musculus
94	315.5	2.4	1496	1	CA25_HUMAN	P05997 homo sapien
95	315	2.4	6632	1	UN89_CABEL	O01761 caenorhabdi
96	314	2.4	1255	1	MUC1_HUMAN	P15941 h mucin 1 p
97	312.5	2.4	1712	1	CA24_HUMAN	P08572 homo sapien
98	311	2.4	2688	1	ZEPI1_MOUSE	Q03172 mus musculus
99	310.5	2.3	3358	1	PGCV_MOUSE	Q62059 mus musculus
100	309	2.3	1087	1	NFH_MOUSE	P19246 mus musculus

CNR2

D NCR2 HUMAN STANDARD; PRT; 2517 AA.

C Q9Y6I8; O00613; Q15416; Q13354; Q9Y5U0;

T 16-OCT-2001 (Rel. 40, Created)

T 16-OCT-2001 (Rel. 40, Last sequence update)

T 16-OCT-2003 (Rel. 42, Last annotation update)

E Nuclear receptor co-repressor 2 (N-COR2) (Silencing mediator of retinoic acid and thyroid hormone receptor) (SMRTE) (Thyroid-, retinoic-acid-receptor-associated co-repressor) (T3 receptor-associated factor) (TRAC) (CTG repeat protein 26).

N NCOR2 OR CTG26.

N Homo sapiens (Human).

S Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

X NCBI_TaxID=9606;

X [1]

X SEQUENCE FROM N.A. (ISOFORM SMRT).

P TISSUE=Pituitary;

P MEDLINE=99178941; PubMed=10077563;

X Ordentlich F., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.; "Unique forms of human and mouse nuclear receptor corepressor SMRT."; Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).

L [2]

N SEQUENCE FROM N.A. (ISOFORM SMRT).

P TISSUE=Cervical adenocarcinoma;

P MEDLINE=9919215; PubMed=10097068;

X Park E.J., Schroen D.J., Yang M., Li H., Li L., Chen J.D.; "SMRte, a silencing mediator for retinoid and thyroid hormone receptor-extended isoform that is more related to the nuclear receptor corepressor."; Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).

R [3]

N SEQUENCE OF 1023-2517 FROM N.A.

P TISSUE=Cervical adenocarcinoma;

X MEDLINE=96008552; PubMed=7566127;

A Chen J.D., Evans R.M.; "A transcriptional co-repressor that interacts with nuclear hormone receptors."; Nature 377:454-457(1995).

R [4]

N SEQUENCE FROM N.A. (ISOFORM TRAC-1).

P TISSUE=Fetal liver;

R MEDLINE=96408715; PubMed=8813722;

R Sande S., Privalsky M.I.; "Identification of TRACs (T3 receptor-associating cofactors), a family of cofactors that associate with, and modulate the activity of, nuclear hormone receptors.";

F Mol. Endocrinol. 10:813-825(1996).

R [5]

N SEQUENCE OF 428-613 FROM N.A.

R TISSUE=Brain cortex;

R MEDLINE=97369492; PubMed=9225980;

R Margolis R.L., Abraham M.R., Gatchell S.B., Li S.-H., Kidwai A.S., Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.; "cDNAs with long CAG trinucleotide repeats from human brain.";

R Hum. Genet. 100:114-122(1997).

R [6]

N INTERACTION WITH MINT.

R MEDLINE=21231190; PubMed=11331609;

R Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C., Hon M., Evans R.M.; "Sharp, an inducible cofactor that integrates nuclear receptor repression and activation";

R Genes Dev. 15:1140-1151(2001).

R [7]

N INTERACTION WITH HDAC10.

R MEDLINE=21839031; PubMed=11739383;

R Fischer D.D., Cai R., Bhatia U., Asselbergs F.A.M., Song C., Terry R., Trojan N., Widmer R., Atadja P., Cohen D.; "Isolation and characterization of a novel class II histone deacetylase, HDAC10.";

R J. Biol. Chem. 277:6656-6666(2002).

R [8]

N Interaction between nuclear receptor activity of some

FT	DOMAIN	1842	1846	POLY-GLY.
FT	DOMAIN	2479	2482	POLY-PRO.
FT	VARSPPLIC	1	1702	Missing (in isoform TRAC-1).
FT	VARSPPLIC	2353	2398	/FTid=Vsp_003412.
FT	VARSPPLIC	2353	2398	Missing (in isoform TRAC-1).
FT	VARSPPLIC	2353	2398	/FTid=VSP_003413.
FT	CONFLICT	7	7	L -> P (IN REF. 2).
FT	CONFLICT	295	295	K -> E (IN REF. 2).
FT	CONFLICT	309	309	L -> W (IN REF. 2).
FT	CONFLICT	352	352	MISSING (IN REF. 2).
FT	CONFLICT	365	365	A -> P (IN REF. 2).
FT	CONFLICT	612	613	SS -> EF (IN REF. 5).
FT	CONFLICT	711	711	S -> T (IN REF. 2).
FT	CONFLICT	724	740	MISSING (IN REF. 2).
FT	CONFLICT	787	796	RTTRAPTEP -> PEDIPAPTES (IN REF. 2).
FT	CONFLICT	804	804	G -> L (IN REF. 2).
FT	CONFLICT	814	814	S -> P (IN REF. 2).
FT	CONFLICT	817	817	A -> S (IN REF. 2).
FT	CONFLICT	889	889	G -> R (IN REF. 2).
FT	CONFLICT	1023	1030	SRSAPPA -> MEAWDAH (IN REF. 3).
FT	CONFLICT	1034	1034	A -> AEKPVFPPA (IN REF. 2).
FT	CONFLICT	1894	1894	K -> T (IN REF. 4).
FT	CONFLICT	2494	2494	P -> A (IN REF. 4).
FT	CONFLICT	2517	2517	AA; 274031 MW; F5805C01761258C0 CRC64;
FT	SEQUENCE	2517	2517	AA; 274031 MW; F5805C01761258C0 CRC64;
Qy	Best Local Similarity	100.0%;	Score 13215;	DB 1; Length 2517;
Qy	Matches 2517;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
Qy	1	MSGSTQLVAQTRATEPRYPHSHLSYPVQIARTHTDVGLLEYQHHSRDYASHLSPGSIQ	60	
Db	1	MSGSTQLVAQTRATEPRYPHSHLSYPVQIARTHTDVGLLEYQHHSRDYASHLSPGSIQ	60	
Qy	61	PORRRPILLSFQPNRSQSLHLRPESHVYLPELGKSEMEFIESKRPRLLEPDLPLRP	120	
Db	61	PORRRPILLSFQPNRSQSLHLRPESHVYLPELGKSEMEFIESKRPRLLEPDLPLRP	120	
Qy	121	SPLLATQCPAGSEDLTKDRSLTGKLEVPSPSPPHTDPELVLPPRLSKEELIQNMDRVD	180	
Db	121	SPLLATQCPAGSEDLTKDRSLTGKLEVPSPSPPHTDPELVLPPRLSKEELIQNMDRVD	180	
Qy	181	REITWVEQOISKLKKQOOLEEBAKAPPEPKVPSPPIESKHSRLVQIIVDENRKKAAE	240	
Db	181	REITWVEQOISKLKKQOOLEEBAKAPPEPKVPSPPIESKHSRLVQIIVDENRKKAAE	240	
Qy	241	AHRLTEGLGQVELPLYNQPSDTRQYHENIKINQAMRKLLTYFKRRNHARKQWKFCQ	300	
Db	241	AHRLTEGLGQVELPLYNQPSDTRQYHENIKINQAMRKLLTYFKRRNHARKQWKFCQ	300	
Qy	301	RYDQLEALEKKVVERIENNPERRAKESKVREYYEKQPEIRKQRELQERMOSRVQGRSG	360	
Db	301	RYDQLEALEKKVVERIENNPERRAKESKVREYYEKQPEIRKQRELQERMOSRVQGRSG	360	
Qy	361	LSWSAARSEHEVSIIDGLSSEQENLEKQMRQLAVIPMLYDADQORIKFINNGLMADPM	420	
Db	361	LSWSAARSEHEVSIIDGLSSEQENLEKQMRQLAVIPMLYDADQORIKFINNGLMADPM	420	
Qy	421	KVYKDRQVMNWSQEQKETPREKFMQHPKNFGLIASFLERTKVAECVLVYLTKKKNYK	480	
Db	421	KVYKDRQVMNWSQEQKETPREKFMQHPKNFGLIASFLERTKVAECVLVYLTKKKNYK	480	
Qy	481	SLVRRSYRRRGKSQQQQQQQQQQQQQQQPMPRSSQBEKDEKEKEAEKEEKEPVEN	540	
Db	481	SLVRRSYRRRGKSQQQQQQQQQQQQQQQPMPRSSQBEKDEKEKEAEKEEKEPVEN	540	
Qy	541	DKEDLLKXKTDGNDENKEAEVASKGRKTANSQGRKRGITRSMANEANSSEAITPQ	600	
Db	541	DKEDLLKXKTDGNDENKEAEVASKGRKTANSQGRKRGITRSMANEANSSEAITPQ	600	
Qy	601	SAELASHELNESSRWTEEMETAKGGLLEHGRNWSAIRMVGSKTVSQCKNFYNYKKRQ	660	
Db	601	SAELASHELNESSRWTEEMETAKGGLLEHGRNWSAIRMVGSKTVSQCKNFYNYKKRQ	660	

Db 1741 PVLVPTGCTPATMDRLAYLPATQPFSSRHSSPLSPGGPHTLTKPTTSSSERDR 1800
 Qy 1801 DRERDREREKSLTSTTTVEHAPIWRPTEOSSSGSGGGSSSRPPASHAHOH 1860
 Db 1801 DRERDREREKSLTSTTTVEHAPIWRPTEOSSSGSGGGSSSRPPASHAHOH 1860
 Qy 1861 SPIGPTQDALQORPSVLHNTGKGIITAVEPSKPTVLRSTSTSSPVRPAATPPATHCP 1920
 Db 1861 SPIGPTQDALQORPSVLHNTGKGIITAVEPSKPTVLRSTSTSSPVRPAATPPATHCP 1920
 Qy 1921 LGGTLGVYPTLMEPVLLEKAPRVARPERPRADTGHAFKAPKAPSGLEPASPCKGSE 1980
 Db 1921 LGGTLGVYPTLMEPVLLEKAPRVARPERPRADTGHAFKAPKAPSGLEPASPCKGSE 1980
 Qy 1981 PRPLVPPVSGHATTARTPAKLAHPHASPPAPASDPHREKTSKPFSTQLELRS 2040
 Db 1981 PRPLVPPVSGHATTARTPAKLAHPHASPPAPASDPHREKTSKPFSTQLELRS 2040
 Qy 2041 LGYHGSSYSPGVEPVSFSLTHDKGLPKHLEELD KSHLEGELRPKQPGPVKLGGEA 2100
 Db 2041 LGYHGSSYSPGVEPVSFSLTHDKGLPKHLEELD KSHLEGELRPKQPGPVKLGGEA 2100
 Qy 2101 AHLPLRLPESQSSPLLQAPGVKHQORVTLAQHISEVITODYTRHHPOQLSAPLP 2160
 Db 2101 AHLPLRLPESQSSPLLQAPGVKHQORVTLAQHISEVITODYTRHHPOQLSAPLP 2160
 Qy 2161 APLYSFPGASCPLDLRRPPSLDLYLPDPDHGAPARGSPHSGGKRSPEPNKTSVLGGED 2220
 Db 2161 APLYSFPGASCPLDLRRPPSLDLYLPDPDHGAPARGSPHSGGKRSPEPNKTSVLGGED 2220
 Qy 2221 GIEPVSPPEGTEPEGHSRAVYPLLYRDGEQTEPSRMGSKPGNTSQPPAFSKLTESNS 2280
 Db 2221 GIEPVSPPEGTEPEGHSRAVYPLLYRDGEQTEPSRMGSKPGNTSQPPAFSKLTESNS 2280
 Qy 2281 AMVSKKQKINKLNTNNEPEYNI SQGTIEFNMPALTGLMTYRSQAVQEHASNTM 2340
 Db 2281 AMVSKKQKINKLNTNNEPEYNI SQGTIEFNMPALTGLMTYRSQAVQEHASNTM 2340
 Qy 2341 GLEAIRKALMGKYDOWESSPLSANAFNPLNASLPAAMPITAADGRSDHITLSPGG 2400
 Db 2341 GLEAIRKALMGKYDOWESSPLSANAFNPLNASLPAAMPITAADGRSDHITLSPGG 2400
 Qy 2401 GKAKVSGRPSRKAKSPAGCLASGDRPPSVSVHSEGCNRTPLTNRWEDRPSAGST 2460
 Db 2401 GKAKVSGRPSRKAKSPAGCLASGDRPPSVSVHSEGCNRTPLTNRWEDRPSAGST 2460
 Qy 2461 PFPYNPLIMELQAGVMASPPPGCLPAGSGPLAGPHAWDEEPKPLICSQYETLSDSE 2517
 Db 2461 PFPYNPLIMELQAGVMASPPPGCLPAGSGPLAGPHAWDEEPKPLICSQYETLSDSE 2517

RESULT 2
 NCR2 MOUSE
 ID NCR2_MOUSE STANDARD; PRT; 2472 AA.
 AC Q9WU42; Q9WU43; Q9WUC1;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Nuclear receptor co-repressor 2 (N-CoR2) (silencing mediator of
 retinoic acid and thyroid hormone receptor) (SMRT) (thyroid-,
 retinoic-acid-receptor-associated co-repressor) (T3 receptor-
 associating factor) (TRAC).
 GN NCR2 OR SMRT.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OC NCBI_taxID=10090;
 RN [1]
 SEQUENCE FROM N.A. (ISOFORMS ALPHA AND BETA).
 RP TISSUE=Brain, and Spleen; PubMed=10077563;
 RX Medline=99178941; PubMed=10077563;
 RA Orthotlich P., Downes M., Xie W., Genin A., Spinner N.B., Evans R.M.;

"Unique forms of human and mouse nuclear receptor corepressor SMRT.";
 Proc. Natl. Acad. Sci. U.S.A. 96:2639-2644(1999).
 [2]
 RN SEQUENCE FROM N.A. (ISOFORM ALPHA).
 RP TISSUE=Embryo;
 MEDLINE=99199215; PubMed=10097068;
 RX Park E.J., Schreen D.J., Yang M., Li H., Li L., Chen J.D.;
 "SMRT, a silencing mediator for retinoid and thyroid hormone
 receptors-extended isoform that is more related to the nuclear
 receptor corepressor.";
 Proc. Natl. Acad. Sci. U.S.A. 96:3519-3524(1999).
 [3]
 RN INTERACTION WITH HDAC7.
 MEDLINE=20107033; PubMed=10640276;
 RX Kao H.-Y., Downes M., Orthotlich P., Evans R.M.;
 "Isolation of a novel histone deacetylase reveals that class I and
 class II deacetylases promote SMRT-mediated repression.";
 Genes Dev. 14:55-66(2000).
 CC -I- FUNCTION: Mediates the transcriptional repression activity of some
 nuclear receptors by promoting chromatin condensation, thus
 preventing access of the basal transcription.
 CC -I- SUBUNIT: Forms a large corepressor complex that contains SIN3A/B
 and histone deacetylases HDAC1 and HDAC2. This complex associates
 with the thyroid (TR) and the retinoid acid receptors (RAR) in the
 absence of ligand, and may stabilize their interaction with TR/RAR.
 CC Interacts with HDAC10 and MINT (By similarity). Interacts with
 HDAC7.
 CC -I- SUBCELLULAR LOCATION: Nuclear.
 CC -I- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Alpha;
 CC IsoId=Q9WU42-1; Sequence=Displayed;
 CC Name=Beta;
 CC IsoId=Q9WU42-2; Sequence=VSP_003414;
 CC -I- TISSUE SPECIFICITY: Ubiquitous. Also widely expressed in early
 embryos.
 CC -I- DOMAIN: The N-terminal region contains repression functions that
 are divided into three independent repression domains (RD1, RD2
 and RD3). The C-terminal region contains the nuclear receptor-
 interacting domains that are divided in two separate interaction
 domains (ID1 and ID2).
 CC -I- DOMAIN: The two interaction domains (ID) contain a conserved
 sequence referred to as the CORNR box. This motif is required and
 sufficient to permit binding to unliganded TR and RAR. Sequences
 flanking the CORNR box determine nuclear hormone receptor
 specificity.
 CC -I- SIMILARITY: Contains 1 SANT-A domain.
 CC -I- SIMILARITY: Contains 1 Myb-like domain.
 CC -I- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF113001; AAD20944.1; -
 EMBL; AF113002; AAD20945.1; -
 EMBL; AF125671; AAD22972.1; -
 TRANSFAC; T04690; -
 MGD; MGI:1337080; Ncor2.
 DR InterPro; IPR001005; Myb_DNA_binding.
 DR Pfam; PF00249; myb_DNA-binding; 2.
 DR SMART; SM00717; SANT; 2.
 DR PROSITE; PS50090; MYB 3; 1.
 DR Nuclear protein; Transcription regulation; DNA-binding; Repressor;
 KW Coiled coil; Alternative splicing
 FT DOMAIN 165 207 COILED COIL (POTENTIAL).
 FT DOMAIN 254 312 INTERACTION WITH SIN3A/B (BY SIMILARITY).
 FT DNA_BIND 429 474 SANT-A (POTENTIAL).

1477 IGSRTFFPHVHLDVMD-ARALERACYE-----ESLKRPGTASSSGSIARGAPVIV 1530
 1418 ITGFSKL--PRGMLEIVPENIKVVERKEDVAGEFVARHTSVVSSGSLVLRST--L 1472
 1531 PELGKPSQSLTYEDHGA-----PFAGHLPRGSPVVTWREPTPLRQEGSLSSSK-ASQDRK 1584
 1473 HEAPKAQLSPGLYDDSGARRTPVSYQNTISKSGSPMNR-----TSDVSSSKSASHERK 1525
 1595 LTSTPRE-----IAKPSHSTVEHHPHIPISPYEHLRLGVSDLYRSHIPLAAPTSPISPRG 1640
 1526 STLTPTQRESIPAKSPVGVNDPIVSH--SPDPHRRSSAAGEVVRSHLPHLTHLP-AMPFH 1582
 1641 IPLDAAAYLPRHLAPNTPHYLPYLYRGYDPTAALN-RQTIINDYITSQOEHNT 1699
 1593 RALDPAAYLRLQRLSPFTGPGVSYQLY-----AMENTRQTIINDYITSQOQVNL 1633
 1700 ATAMAQADMLRGLSPRESSALANYAAGFRGIIDLSQVPHLPVLVPTPGTPTAMDRLA 1759
 1634 -----REDVTRGLSPREQLGLPYPA-TRGIIDLTNMP-TELPHAGGTSTPFMDRIT 1685
 1760 YLTPAQPFSSR-HSSSPLSGGPHLTKPTTSSSERERDRDRDRERERESILTST 1818
 1686 YTPGTQVTFPPPPYNAASLSPGPHLTHL-----AAAASERERERERERERERERER 1742
 1819 TTVEHAP---IWRPGTEQSSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 1874
 1743 ERIAAPADLYLRPGSEQ-----PGRPGSHGVVRSPSP-SVRTQETILQOR 1787
 1875 PSVLNHTGMKGIITAVEPSKPTVLIRSTSTSPV-----RPAATFPATHCPGLGGTLDGVP 1930
 1788 PSVFQGTNGTSVITLDPDPAQRLMPLSPGSGSISQGLPAGRYNTAADA-LAALVDAAS 1846
 1931 TLMPEVLLKEAPR-----VARPERPRADTGHAFLAKPPARSPGLEPASSPSK 1977
 1847 APOMDVSKTESKHEARLEENLRSSAAVSQQLKQKNLEVEKRSVQCVCTSSALPSG 1906
 1978 GSEPRELV-----PPVSGHATITARTPAKN--LAPHASDPDPAPPASADPHRE 2024
 1907 KAQPHASVVVSEAGKDKGPPKSRVEELRTGKTTITAAFNFDVIIITQIASDKARER 1966
 2025 KTSQKFSIQELRLSLGVHSGSYSPGVEPVSSPSLTHDKGLPKHLELDKSHLEG 2084
 1967 GSQSSDS-----SSLSHRYETASDAIEVISPASSPAPQEKQAYQVDMVKANQAE 2020
 2085 ELRKPQPGPKLGGEAHLPLR-----PLPSQSPSSPLLQ--APGVKHQVTVLTAQ 2137
 2021 ESTROYEGP-----LHHVRSQESPSPOQPPPLPSSQSGMGQVPRTHRLITLAD 2071
 2138 HISEVITQDYTRH-POOLSAPL-----PAPLYSFGACSPVLDLRRPPSDLYLPPPD--- 2189
 2072 HICQITQDFARNQVPSQASTSTFTQSPSALSSTP-----VRKTSRYSPPSQSOT 2123
 2190 --HGAPA-RGSPH-----SEGGKRSPEPNKTSVLGGEDGIEPVSPPEGMTEFG-HSRSA 2240
 2124 VLHPRGPRVSPENLVDKSRGSRGKSPERSHI---PSEYEPISPPQ---PAVHEKOD 2177
 2241 VYPLLYRQDQTEPSRMGKSPGNTSPAPFAFKLTESNSAMVSKKQKINKLNTNHN 2300
 2178 SMLLSQGVDPAPQRSDRSRSPGISYLPFFTKL-ESTSPMWKSKQKIFRKINSGGG 2236
 2301 EPEYNISQGTIEFNMPAITGTGLMVTYSQAOVGHASTNMGLEAIRKALMGKYDQWEE- 2359
 2237 DSDMAAQPGTEIENLPAVITSGVSRSHSFADPAS-NLGLIEDIIRKALMGSDFKVED 2295
 2360 -----SPPLSANAFNPLNASASLPAAMPITTAAGRSHTLTPCGG-GRKAKVGRSPSRK 2413
 2296 HGVVMSHVV---GIMPGSASTSV-----VTSSEARDEGEPSPHAGVCKPKLINKNSRK 2347
 2414 AKSPAPGLA--SGDRPPSVSVSHSEGCNRRTPLTNVRWEDRPSAGSTFPFYNPLIMRL 2471
 2348 SKSPIPGOSYLGTERPSSVSVSHSEGDYHROTP--GNAWEDRPSSTGSTQFPYNPLTIRM 2405

2472 QAGVMASPPPLGAPAGSPL--AGPH---HAWDEEPKPLLCQSYETLSDSE 2517
 2406 ----LSSTPTQIACAPSAITQAAPHQONRIWEREPALLSAQYETLSDSD 2452

RESULT 4
 NCRL_HUMAN
 ID NCRL_HUMAN STANDARD; PRT; 2440 AA.
 AC O75376; Q9UPV5; Q9UQ18;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Nuclear receptor co-repressor 1 (N-COR1) (N-COR).
 GN NCOR1 OR KIAA1047.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Fetal brain;
 RX MEDLINE=98393736; PubMed=9724795;
 RA Wang J., Hoshino T., Redner R.L., Kajigaya S., Liu J.M.,
 RT "ETO, fusion partner in t(8;21) acute myeloid leukemia, represses
 RT transcription by interaction with the human N-COR1/HDAC1
 RT complex.";
 RL Proc. Natl. Acad. Sci. U.S.A. 95:10860-10865(1998).
 RN [2]
 RP SEQUENCE OF 782-2440 FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=99397452; PubMed=10470851;
 RA Kikuno R., Nagase T., Ishikawa K.-I., Hirose M., Miyajima N.,
 RA Tanaka A., Kotani H., Nomura N., Ohara O.;
 RT "Prediction of the coding sequences of unidentified human genes. XIV.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL DNA Res. 6:197-205(1999).
 RN [3]
 RP SEQUENCE OF 974-2440 FROM N.A.
 RX MEDLINE=99375328; PubMed=10444336;
 RA Nagaya T., Chen K.-S., Fujieda M., Ohmori S., Richer J.K.,
 RA Horwitz K.B., Lupski J.R., Seo H.;
 RT "Localization of the human nuclear receptor co-repressor (hN-COR) gene
 RT between the CMT1A and the SMS critical regions of chromosome
 RT 17p11.2.";
 RL Genomics 59:339-341(1999).
 CC -!- FUNCTION: Mediates the transcriptional repression activity of some
 CC nuclear receptors by promoting chromatin condensation, thus
 CC preventing access of the basal transcription.
 CC -!- SUBUNIT: Interacts with HDAC7 (By similarity). Forms a large
 CC corepressor complex that contains SIN3A/B and histone deacetylases
 CC HDAC1 and HDAC2. This complex associates with the thyroid (TR) and
 CC the retinoid acid receptors (RAR) in the absence of ligand.
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
 CC -!- DOMAIN: The N-terminal region contains repression functions that
 CC are divided into three independent repression domains (RD1, RD2
 CC and RD3). The C-terminal region contains the nuclear receptor-
 CC interacting domains that are divided in two separate interaction
 CC domains (ID1 and ID2).
 CC -!- DOMAIN: The two interaction domains (ID) contain a conserved
 CC sequence referred to as the CORNR box. This motif is required and
 CC sufficient to permit binding to unliganded TR and RARs. Sequences
 CC flanking the CORNR box determine nuclear hormone receptor
 CC specificity.
 CC -!- SIMILARITY: Contains 1 SANT-A domain.
 CC -!- SIMILARITY: Contains 1 Myb-like domain.
 CC -!- SIMILARITY: BELONGS TO THE N-COR NUCLEAR RECEPTOR COREPRESSORS
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way


```
FT CONFLICT 484 484 R -> W (IN REF. 2).
FT CONFLICT 497 497 A -> V (IN REF. 2).
SQ SEQUENCE 533 AA; 57794 MW; 7DF60F8228227EC2 CRC64;

Query Match
Best Local Similarity 5.4%; Score 714.5; DB 1; Length 533;
Matches 205; Conservative 68; Mismatches 192; Indels 67; Gaps 24;

QY 2016 ASADPHREKTKQSFISQIELESLRSLGHGSSYSPEGVPPVSPVSLTHDKGLPKHLE 2075
Db 38 ASDKADREKGSQSDSS-----SSLSSHRYEAPSDAIEVISPASSAPPQKEKPTQYQE 91
QY 2076 ELDKSHLEGELPKQFGPKVLGCEAAHPLHR-----PLPESQPSSSLPLQTAPG---VK 2127
Db 92 MYKANQAEKESQYEGP-----LTHYSQQGSPSPQQPPLPSSQ-AEGMGQVP 141
QY 2128 GQHVVTTLAQHISEVITDTRHHQQQLSAPLAPLYSPFGA--SCPVLDRRPPSDLYL 2185
Db 142 RTHRLITLADHLCIQLITQDFARN--QVPSQPTSTFTQSPSALSTPV---RTKPSRYS 196
QY 2186 PPPD-----HGAPA-RGSPH-----SEGKRSPEPNKTSVLGGGEGIEPVSPPEGMTPE 2234
Db 197 PESQSQTVLHPRGPRVSPENLVKSRGSRPGKSPERSHI---PSEPYEPTSPQG---P 250
QY 2235 G-HRSRAVPLLYRGEOTEPSRMGSKGPNQTSQPPAFSKLTSNSAMVSKKQEIINK 2293
Db 251 AVHERQDSMLLISQRGMDPAEQRSRSPGSIYLPYFTKL-ESTSPMVKSKQEIIRK 309
QY 2294 LINTNRNPEYNISQGTIEFNMPAITGTGLMYSQAOVHASTMGLAIIIRKALWGK 2353
Db 310 LNSGGGSDMAAQAQGTIEFNLPVTTSGAVSSRSHPADPAS--NLGLEDIIIRKALWGS 368
QY 2354 Y-DOWEEPPPLSANAFNPLNASASLPAAMPITAADGRSDHTLTSPGGG-GRKAVGRPSS 2411
Db 369 PDDKVEDHGVMPHPVGVVPGSASTSV---VTSSSTRDEGDPSPHSGVCKPKLINKNS 425
QY 2412 RKAQSPAPG--LASGDRPPSVSVHSEGDGNNRRTPLTNVMDRPPSSAGSTPPFYNPLIM 2469
Db 426 RKKSPIPGQNYLGTGRTSSSVSVHSEGDYHRTPE--GWAMEDRPPSTGSTQFPYNPLTI 483
QY 2470 RLQAGVMAAPPPGGLPACSG-PLAGPH---HAWDEEPKLLCSQYETLSDSE 2517
Db 484 RM---LSSTPTPTTACAPSAITQAAAPHQOSRIWEREPAPLLISAQYETLSDSD 532
```

RESULT 6

```
MINT_MOUSE
ID MINT_MOUSE STANDARD; PRT; 3644 AA.
AC Q62504; Q80TN9; Q99PS4; Q9QZM2;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
GN MINT OR SHARP OR KIAA0929.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE OF 49-3644 FROM N.A. (ISOFORM 1), FUNCTION, SUBCELLULAR
RP LOCATION, TISSUE SPECIFICITY, DNA-BINDING, AND INTERACTION WITH MSX2.
RC TISSUE-Testis;
RX MEDLINE=93379811; PubMed=10451362;
RA Newberry E.P., Latifi T., Towler D.A.;
RT "The RRM domain of MINT, a novel mxx2 binding protein, recognizes and
RT regulates the rat osteocalcin promoter.";
RL Biochemistry 38:10678-10690(1999).
[2]
RN SEQUENCE OF 1-112 FROM N.A.
RA Adachi J., Aizawa K., Akimura T., Arakawa T., Carninci P., Fukuda S.,
RA Hashizume W., Hayashida K., Hirozane T., Hori F., Imotani K.,
RA Ishii Y., Itoh M., Kagawa I., Kawai J., Kojima Y., Kondo S., Konno H.,
RA Koya S., Miyazaki A., Murata M., Nakamura M., Nomura K., Numazaki R.,
```

```
RA Ohno M., Ohsato N., Saito R., Sakazume N., Sano H., Sasaki D.,
RA Sato K., Shibata K., Shiraki T., Tagami M., Takeda Y., Waki K.,
RA Watahiki A., Muramatsu M., Hayashizaki Y.;
RL Submitted (OCT-2002) to the EMBL/GenBank/DBSJ databases.
[3]
RN SEQUENCE OF 69-3644 FROM N.A. (ISOFORM 2), AND VARIANTS THR-348;
RP PHE-762; PHE-773 AND LEU-933.
RC STRAIN=ICR; TISSUE=Brain;
RA Sakamoto T., Gotou T., Isagawa Y., Mimura H., Kimura K., Kawaichi M.;
RT "MINT/spen negatively regulates Notch signaling by inhibiting RBP-
RT J/Su(H) activity.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBSJ databases.
[4]
RN SEQUENCE OF 318-578 FROM N.A.
RP TISSUE=Cochlea;
RX MEDLINE=97237053; PubMed=9119401;
RA Crozet P., El-Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.,
RA Hamet C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,
RA Weil D., Fujol R., Petit C.;
RT "Cloning of the genes encoding two murine and human cochlear
RT unconventional type I myosins.";
RL Genomics 40:332-341(1997).
[5]
RN SEQUENCE OF 2598-3644 FROM N.A.
RP TISSUE=Brain;
RX MEDLINE=22579291; PubMed=12693553;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Aizawa H., Yuasa S.,
RA Nakajima D., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT II. The complete nucleotide sequences of 400 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:35-48(2003).
[6]
RN TISSUE SPECIFICITY.
RP MEDLINE=22261914; PubMed=12374742;
RA Oswald F., Kostecka U., Astrahantseff K., Bourteele S., Dillinger K.,
RA Zechner R.M., Wilda M., Hameister H., Knoeschel W., Liptay S.,
RA Schmid R.M.;
RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
RT pathway.";
RL EMBO J. 21:5417-5426(2002).
[7]
RN FUNCTION, AND TISSUE SPECIFICITY.
RX MEDLINE=22483652; PubMed=12594956;
RA Kuroda K., Han H., Tani S., Tanigaki K., Tun T., Furukawa T.,
RA Taniguchi Y., Kurooka H., Hamada Y., Toyokuni S., Honjo T.;
RT "Regulation of marginal zone B cell development by MINT, a suppressor
RL of Notch/RBP-J signaling pathway.";
IMMUNITY 18:301-312(2003).
CC -!- FUNCTION: Essential corepressor protein, which probably regulates
CC different key pathways such as the Notch pathway. Negative
CC regulator of the Notch pathway via its interaction with RBPSUH,
CC which prevents the association between NOTCH1 and RBPSUH, and
CC therefore suppresses the transactivation activity of Notch
CC signaling. Blocks the differentiation of precursor B cells into
CC marginal zone B cells. Probably represses transcription via the
CC recruitment of large complexes containing histone deacetylase
CC proteins. May bind both to DNA and RNA.
CC -!- SUBUNIT: Interacts with NCOR2, HDAC1, HDAC2, RBBP4, MED3 and
CC MYTALL1. Interacts with the nuclear receptors RAR and PPARG.
CC Interacts with RAR in absence of ligand. Bind to the steroid
CC receptor RNA coactivator SRA (By similarity). Interacts with MSX2.
CC Interacts with RBPSUH; this interaction may prevent the
CC interaction between RBPSUH and NOTCH1.
CC -!- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q62504-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q62504-2; Sequence=VSP_008564;
CC Note=No experimental confirmation available;
```


QY 2409 PSSRKAKSPAGLASGRPP-----SYSSVHSEGDGNNRTPLN-RVWEDRPPSAGSTPF 2462
 DB 3247 PSSQLQGLPL-----TPPVVTHGVQIVHSSGELFQEVRYGVDVRYTHAPAOQLTHTQF 3299

QY 2463 PY-NPLMLRQAGWASPPPPGLPAGSGPLAGP 2494
 DB 3300 PVASSISLASRTKTSQAQVPEGEPLQSTQSAQP 3332

RESULT 7

MINT HUMAN STANDARD; PRT: 3664 AA.

AC Q96T58; Q9H9A8; Q9NWH5; Q9UQ01; Q9Y556;
 DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Mx2-interacting protein (SMART/HDAC1 associated repressor protein).
 GN MINT OR SHARP OR KIAA0929.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1] SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, INDUCTION,
 RP RNA-BINDING, AND INTERACTION WITH NCOR2; HDAC1; HDAC2; RBBP4; MBD3;
 RP RAR AND MTA1L1.
 RC TISSUE=Liver, and Pituitary;
 RX MEDLINE=12311190; PubMed=11331609;
 RA Shi Y., Downes M., Xie W., Kao H.-Y., Ordentlich P., Tsai C.-C.,
 RA Hon M., Evans R.M.;
 RA "Sharp, an inducible cofactor that integrates nuclear receptor
 RT repression and activation.";
 RL Genes Dev. 15:1140-1151(2001).
 RN [2] SEQUENCE FROM N.A.
 RP Bird C.;
 RL Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.
 RN [3] SEQUENCE OF 294-3664 FROM N.A.
 RA Rhodes S., Huckle E.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
 RN [4] SEQUENCE OF 793-1595 FROM N.A., AND VARIANT PRO-1091.
 RC TISSUE=Embryo, and Teratocarcinoma;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
 RA Tanase T., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K.,
 RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
 RA Wakamatsu A., Nakamura Y., Nagahara K., Masuho Y., Oshima A.;
 RT "NEDO human cDNA sequencing project.";
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
 RN [5] SEQUENCE OF 2002-3664 FROM N.A.
 RP TISSUE=Brain;
 RX MEDLINE=9246063; PubMed=10231032;
 RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirose M.,
 RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
 RA "Prediction of the coding sequences of unidentified human genes. XIII.
 RT The complete sequences of 100 new cDNA clones from brain which code
 RT for large proteins in vitro.";
 RL Nucleic Acids Res. 27:1213-1222(1999).
 RN [6] INTERACTION WITH PP4R.
 RP MEDLINE=21874127; PubMed=11867749;
 RX Shi Y., Hon M., Evans R.M.;
 RA "The peroxisome proliferator-activated receptor delta, an integrator
 RT of transcriptional repression and nuclear receptor signaling.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:2613-2618(2002).
 RN [7] FUNCTION, AND INTERACTION WITH RBP4.
 RP MEDLINE=22261914; PubMed=12374742;
 RX Oswald F., Kostezka U., Astrahantseff K., Bourteelle S., Dillinger K.,

RA Zechner U., Ludwig L., Wilda M., Hameister H., Knoechel W., Liptay S.,
 RA Schmid R.M.;
 RT "SHARP is a novel component of the Notch/RBP-Jkappa signalling
 RT pathway.";
 RL EMBO J. 21:5417-5426(2002).
 RN [8] X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF SPOC DOMAIN.
 RP MEDLINE=22777836; PubMed=12897056;
 RA Ariyoshi M., Schwabe J.W.R.;
 RA "A conserved structural motif reveals the essential transcriptional
 RT repression function of Spen proteins and their role in developmental
 RT signaling.";
 RL Genes Dev. 17:1909-1920(2003).
 CC -1- FUNCTION: Essential corepressor protein, which probably regulates
 CC different key pathways such as the Notch pathway. Negative
 CC regulator of the Notch pathway via its interaction with RBP4, and
 CC which prevents the association between NOTCH1 and RBP4, and
 CC therefore suppresses the transactivation activity of Notch
 CC signaling. Blocks the differentiation of precursor B cells into
 CC marginal zone B cells. Probably represses transcription via the
 CC recruitment of large complexes containing histone deacetylase
 CC proteins. May bind both to DNA and RNA.
 CC -1- SUBUNIT: Interacts with MSX2 (By similarity). Interacts with
 CC NCOR2, HDAC1, HDAC2, RBBP4, MBD3 and MTA1L1. Interacts with
 CC RBP4; this interaction may prevent the interaction between
 CC RBP4 and NOTCH1. Interacts with the nuclear receptors RAR and
 CC PP4R. Interacts with RAR in absence of ligand. Bind to the
 CC steroid receptor RNA coactivator SRA.
 CC -1- SUBCELLULAR LOCATION: Nuclear. Associates with chromatin.
 CC -1- TISSUE SPECIFICITY: Expressed at high level in brain, testis,
 CC spleen and thymus. Expressed at intermediate level in kidney,
 CC liver, mammary gland and skin.
 CC -1- INDUCTION: By hormone 17-beta-estradiol (E2).
 CC -1- DOMAIN: The RID domain mediates the interaction with nuclear
 CC receptors (By similarity).
 CC -1- DOMAIN: The SPOC domain, which mediates the interaction with
 CC NCOR2, is essential for the repressive activity.
 CC -1- SIMILARITY: Belongs to the Spen family.
 CC -1- SIMILARITY: Contains 1 RID (receptor interacting) domain.
 CC -1- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
 CC -1- SIMILARITY: Contains 1 SPOC domain.
 CC -1- CAUTION: Ref.2 sequences differ from that shown due to erroneous
 CC gene model prediction.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AF356524; AAK52750.1; -
 CC EMBL; AL034555; CAB85442.1; ALT SEQ.
 CC EMBL; AL034555; CAB85444.1; ALT SEQ.
 CC EMBL; AL450998; -; NOT ANNOTATED CDS.
 CC EMBL; AL036858; CAB51072.1; ALT_INIT.
 CC EMBL; AK000882; BAA91405.1; ALT_INIT.
 CC EMBL; AK022949; BAB14324.1; ALT_INIT.
 CC EMBL; AB023146; BAA76773.1; -
 CC InterPro; IPR000504; RNA_rec_mot.
 CC PDB; 1OW1; 19-AUG-03.
 CC Pfam; PF00076; rtm; 4.
 CC SMART; SM00360; RRM; 4.
 CC PROSITE; PS50102; RRM; 4.
 CC PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
 CC PROSITE; PS50917; SPOC; 1.
 CC Transcription regulation; Coiled coil; 3D-structure; Polymorphism.
 KW Repeat; RNA-binding; Coiled coil; 3D-structure; Polymorphism.
 FT DOMAIN 1 573 DNA-BINDING (BY SIMILARITY).
 FT DOMAIN 6 81 RNA-BINDING (RRM) 1.
 FT DOMAIN 335 415 RNA-BINDING (RRM) 2.
 FT DOMAIN 438 513 RNA-BINDING (RRM) 3.

FT	DOMAIN	517	589	RNA-BINDING (RBM) 4.		Db	1635	PSVGPSPV-----TVTVLESAP-----SALEKTTGDKTV-----1666
FT	DOMAIN	688	715	COILED COIL (POTENTIAL).		Qy	654	FNKKRQNLDELQOHLKMEKERNARRKKKAPAAASEAAAPPVVVEDEMEASGVSGN 713
FT	DOMAIN	977	1004	COILED COIL (POTENTIAL).		Db	1664	-----EALPVTEKTVEPATVSEAKPASEPAPA 1692
FT	DOMAIN	1170	1191	COILED COIL (POTENTIAL).		Qy	714	BEEMVEEABALHASGNEVPRGECSPATVNNNSDTEISIPGPHTEBAAKDXTGONGKPKPATL 773
FT	DOMAIN	1408	1428	COILED COIL (POTENTIAL).		Db	1693	PVEQLEQV-----DLPPGA-----DPDKAAMMP---AGVEEGSSGDQPP-YL 1731
FT	DOMAIN	1496	1529	COILED COIL (POTENTIAL).		Qy	774	GADGPPPPPTPPRTSRAP--IBPTPASEATGATPPPPPPSPAPPPVPPVPEKEEET 831
FT	DOMAIN	1592	1612	COILED COIL (POTENTIAL).		Db	1732	DAK-----PPTPGASFQAESNVDPEDS-----TQPLSKPAQKSEANEPKAEKPDAT 1780
FT	DOMAIN	1928	1944	COILED COIL (POTENTIAL).		Qy	832	AAAPP-----VEEGEOKPPAAAEALAVD---TGKAERPVKSECTEEAEEGPAGKDAEAA 883
FT	DOMAIN	2201	2707	RID.		Db	1781	ADAEFDANQKAEAPESQPPASEDELDVDPVAAKDKKPKNSKSKTVPQAAANVIVEKPV 1840
FT	DOMAIN	3498	3664	SPC.		Qy	884	EATAEGALKAEKKEGGSGRATTAKSSGAPOQSDS--SATCSADEVDEAEGDGKRLLSRPP 942
FT	DOMAIN	2130	2464	INTERACTION WITH MSX2 (BY SIMILARITY).		Db	1841	TRKSERIDREXLRKRSNPRGEAQKLELMEAEKTIKTASKNSAADLE-----HPEP 1892
FT	DOMAIN	2709	2870	INTERACTION WITH RBPSUH (BY SIMILARITY).		Qy	943	SL-LTPT-----GDPRANSPQK-PLD-----LKQLQKRAAAPIPIQVTK 980
FT	DOMAIN	125	277	ARG-RICH.		Db	1893	SLPLSRTRRRNRSVYATMGD--HENRSPVKEPVQPRVTRKRLEREQEAAPV--TTPR 1949
FT	DOMAIN	240	325	SER-RICH.		Qy	981	VHEPR-----EDAAPTKPAAPP-----PQNL-----QPESDAP 1012
FT	DOMAIN	616	810	ARG-RICH.		Db	1950	RGRPKTRRRADEENEAQKPAETLKPPEGWRSPRQKTAAGGGPGKKGKNEPKVDAT 2009
FT	DOMAIN	624	697	TVR-RICH.		Qy	1013	QOQGSPPRG-----KSRSPAPPADKEAFAPAAEAQKLPDPPCWTSGLPFPVPPREVIKASP 1067
FT	DOMAIN	2428	2520	PRO-RICH.		Db	2010	REPATEVGPQIGVKESMEFKAABEAGSGEQKDRKD--AGTDKNPETAPEVVEVKKP 2067
FT	DOMAIN	3220	3482	PRO-RICH.		Qy	1068	HAPDSAFSYAPPGHPLPLGLHDTARVLP RPPTISNPPPLISAKHPSVLRIQIGAISQ 1127
FT	VARIANT	970	970	A -> V (in dbSNP: 848208).		Db	2068	APENKSKS-----KRGSRNSRLAVDKSASLKN-----VDAAVSFRGAAQAQAGERES 2114
FT	VARIANT	1091	1091	/FTId=VAR 017119.		Qy	1128	GMSVOLHPVYSEHAKAPVGTMTGLPLPMDP-----KKLAPFSG-----VKOEOL-- 1172
FT	VARIANT	2360	2360	L -> P (in dbSNP: 848209).		Db	2115	GV-VAVSEKSESPOKEDGLSSQLKSDPVDPKPEKEDVASGSPSEATQLAKQMELEQ 2173
FT	CONFLICT	956	956	G -> D (in REF. 4).		Qy	1173	-----SPRQAGPESGLVPTAQAEASVLRGTLGVS-----1203
FT	SEQUENCE	3664	3664	AA; 402245 MW; 5228C58533B5B27B CRC64;		Db	2174	AVEHTAKLAEASASAAVKADAEPLA--PEDRDKPAHQASETELAALAAIGSIINDISGEPE 2231
Query Match				4.3%; Score 569.5; DB 1; Length 3664;		Qy	1204	-----PGSITKGPSTRV-----PDSASITVGRSITHTGTPADVLYKGTI--TRIIGED 1250
Best Local Similarity				19.6%; Pred. No. 3.4e-11;		Db	2232	NFPAPPPYPGESQTDLPQAGAAQALQFSE-----EGMETDEAVSGILETEATES 2281
Matches				568; Conservative 349; Mismatches 973; Indels 1011; Gaps 137;		Qy	1251	S-----PSRLDRGREDSLPKGHVITYEKGKGVLSYEGGMSVTQCSKEDGRSSS 1298
71	EFQGNERSQELHLRPE-----SHSYLPELGKS-----EMEFIESKR---PRLELLP 114					Db	2282	SRPPVNAPODAGPTDTTKAARGNSSETSHSVPEANGSK-----EVEVLVRKDKGKQ-- 2333
1142	ERKSQGEKSHSVNTEKIGIDIDHTQSYRKQWESRRKQOMEMETAKSEKFGSPKDV-- 1199					Qy	1299	GPPHETAAPKRTYDMMEGRVGRVIASSAIEGLMGRAIPPRPHSPHHLKEQHIRG--SI 1355
115	DLPLRSPILATQPGAGSBDLTQKSLTGKLEBPVSPSPPHPTDPELELVPRLSKEBELIQ 174					Db	2334	-----KTYRSRRKENTNKV-----VAPVESHVP-----ESNQAQGESPA 2369
1200	DEYERRSLVHEVGKP--PQDVTD-----SPPSKK-----K 1228					Qy	1356	TGIPRPSVVEAQEDYLREAKLLKREGTPPPPPPSRDLTBAYTKQALGPLKPAHEGLV 1415
175	NMDRVDREI--TWVEQIQSKLKKKQOQLEEAAKPP-----EPEKPVSPPIES----- 221					Db	2370	NEGTTVQHPEAPQ-----EEKQSEKPHSTPPQSCSDLSKIPSTE-----2409
1229	RMHDVDFDICTRERNYRSRQISDSEKTSRGSPSVRHGSHFEDDPTGSPRLLSVKGSP 1288					Qy	1416	ATVKEAGRSIHEIPREELRHT-----PELPLAPPLKEGSTTQGTPLKYDTGASTTGSKK 1470
222	-----KHRSILVQIYIDENRKK--AEAARHILEGLGPQVELPLYNQPSDTRQ 265					Db	2410	-----NSQEIISVEERTPTKASVPDLPPPPQ-----APVDEEPQA-----R 2447
1289	KVDEKVLPSYNTVREESLKFPYDSSRRQWADWAKIKLSVLNSEDENRW-----DSQM 1344					Qy	1471	HDVRLSIGSPKRTFP--PVHPLDVNADARALERACYEESLKRSGPTASSGGSGTARGAPV 1528
266	YHENIKINQAMRKLILYFKRRNHARKWKQKFCORYDOLMEALEKKVRIENNPR-- 322					Db	2448	FRVHSIIESDPVTPPSDPSIPIPLPSV-----TAAKLSPPVASC-- 2487
1345	KQDAGRFDVFPNSII-----KEDSLRKSVRDL--EPGEVPSDSDEGEHKSHPRASAL 1398					Qy	1529	IVPELGKPRQSP-----LTYEDHGAPFAGHLPRGSPVTWREPTPLRQBSLSS--SK 1578
323	-----RAKESKVREYKQPEIKQRELQERQSVRGQSGSL-----SMSAAR 367					Db	2488	-----GIHQSPPTKVTETWITROE-----EP

QY	1579	ASQDKLTSPRETAKS-----PHSTVPEHHPHPISPYEHLLRGVSGVDLYR	1625
Db	2531	SSTURKILMDKYVSATSVSTSTVTTAAEPVSAAPCLHEAPPPVD-----SKKPLEE	2584
QY	1626	SHIFLAEDFTSIPRGIFLDAAYLPHRIAPNTYPLHYPPVILRGYPTALENROTI	1685
Db	2585	KTAFPVTTNNSBIQASEVLVAADKEKVAPIAPKIT-----SVISMPVSIIDLENSQKI	2637
QY	1686	INDVITSQOMHNTATAWAQADMLRGLSPRESSLALNYAAGPRGIIDLQVPHLPVLVP	1745
Db	2638	-----TLAKAPQTLTGL-----VSALTGLVNVSLVP-VNALKG	2670
QY	1746	PTPGPATMDRLAYLPTAPQPFSSRHSSPLSPGPHLTHTKPTTSSSSRERDRDRD	1805
Db	2671	PVKGSVTLKSLVS-----TPAGPVNVLKGPV-----	2697
QY	1806	RDREKSIILSTTTTVEHAPI-----WRPGTEQSSGS-----SGSS	1841
Db	2698	-----NVLTGCPVNVLTTPVNAVGTVNAAPGTVNAASAVNATASAVTAVTAVTAAAS	2750
QY	1842	GGGGS-----SSRPASHSHAHQHSPIPRTOALQOQRPVSLHNTGMKGIIIT	1888
Db	2751	GGVTATTGTVTMAGAVIAPSTKCKORASANENSRHFPGSMFVIDDRA-----DAG-SGAGL	2806
QY	1889	AVEPSKPTVLRSTS---TSSPVPRPAA---TPPPATHCLPGTLDGVYPTLMEPVLLPKEA	1942
Db	2807	RVNTSEGVLLSYSGOKTEGPORISAKISQIPPAS-----AMDIEFOOSVSKSQVKPDS	2860
QY	1943	PRVARP---ERPADTGHFLA-----KPPARSGLEPASSPSKSEPRPL---	1984
Db	2861	VTASQPPSGQPAQAYANVATHSLVLTATQYNASPVISVK-ADRFSL-EKEPEIHLS	2918
QY	1985	-----VPPVSGHATARTPA-----	1999
Db	2919	VSTPVTQGTGVKVLTOGINTPVLVHNLQVLTPSIVTNNKLDAPVTLKTIETKVLQPANL	2978
QY	2000	-XNLAPHASDPAPPAPASADP---HREKTQKFP---STQELSLGLVGHSSYSPEGEV	2054
Db	2979	GSTLTTPHH---PPALPSKLTPEVNVHVPSPGSIADRTVSHLAAAKLDAHSRPSGPGPS	3034
QY	2055	PVSPVSPSLTHDKGL-----PKH---LEELDKSHL- 2082	
Db	3035	SPRASHPASTASTALSTWATVMAAGIPVPQFISSIHPEQSVIMPPHISITQVLSHLS	3094
QY	2083	EGELRPKQPG-----PVLGGEAAHLPHLRPLPESQSSPLIQTAPGVKQHVTVTLAQH	2138
Db	3095	QGEVRMNTPLPSITYSIRPEALHSR-APL---QP-----QQIEVRA--- 3133	
QY	2139	ISEVITQDYTRHHPOQLSAPLAPLYSFFGASCPCVLDLRRPPSD---LYLPPPDHGAPAR	2195
Db	3134	-----PQRASTPQAP-----AGVPALASQHPPEEVHYHLPVARATAPVQ	3174
QY	2196	GS-----PH-----SEGGRSPPEPKNTSVLGGEGDG 2221	
Db	3175	SEVLVMQSEYRLHPYTPVRDVRIMVHPHTVAVSEQPRAADGVVKKVPFASKAP---QOPG	3230
QY	2222	IEPVSPPEGTEPCHRSAY-----YPLLXRDGEQTEPSRMGSKSPGNTSO-----PP	2269
Db	3231	KEAAKTDPKAAETPTTAPYVPVPLPAPAPAPHGE-----ARLTVTPSNQLOGLPLTTP	3286
QY	2270	APFSKLTESNAMYKSKQBINKLNTNHRNE---PEYNISQPGTEIFNMPAITGTGLMTY	2327
Db	3287	-----VVVTHGVQI-----VHSSGELFQEVRYGD-----IRTV 3314	
QY	2328	RSQAVQEH-----ASTNNGLEAIRKALMGYDQWEE---SPLSANAFNPLNASASLPAAMP	2382
Db	3315	HPPAQLTHIQPFAASVGLFSRKTAAQAGPPPEGEPLQPPQVOSTQAPQAPCPPSQ- 3373	
QY	2383	ITAADGRSDHLLTSPGGGKAK-----VSGRPSSRRKAKSPAPGLASG---DRPP-----SVSSV	2433
Db	3374	-----LGQGGQPPSSKMPQVQAEKGTGTGVEQPRLPAGPANRPPEPHTQVORA	3422

QY	2434	HSE-GDCNRRRTPLTRVNWEDRPFSSAGSTPPFFNPLIMELQAGVMASPPPPGLRAGSGELA	2492
Db	3423	QRETOGTFPFPSPVSMKPDLPVSLPTQTAPKQPLFVPTTSG-----PSTPPG---LV	3472
QY	2493	GFPHAWDEEPPK-----PLICSQ 2509	
Db	3473	LPHTFQAPKQDSSPPLTSTQ 3493	

RESULT 8

BAT2_HUMAN

ID BAT2_HUMAN STANDARD; PRT: 2142 AA.

AC P48634;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Large proline-rich protein BAT2 (HLA-B-associated transcript 2) (G2).

GN BAT2 OR G2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID:9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=T-cell;

RX MEDLINE=90192810; PubMed=2156268;

RA Banerji J., Sands J., Strominger J.L., Spies T.;

RT "A gene pair from the human major histocompatibility complex encodes

RT large proline-rich proteins with multiple repeated motifs and a

RT single ubiquitin-like domain.";

RL Proc. Natl. Acad. Sci. U.S.A. 87:2374-2378(1990).

RN [2]

RP SEQUENCE OF 1-1860 FROM N.A.

RX MEDLINE=93272029; PubMed=8499947;

RA Iris F.J.M., Bougueleret L., Prieur S., Caterina D., Primas G.,

RA Perrot V., Jurka J., Rodriguez-Tome P., Claverie J.-M., Dausset J.,

RA Cohen D.;

RT "Dense Alu clustering and a potential new member of the NF kappa B

RT family within a 90 kilobase HLA class III segment.";

RL Nat. Genet. 3:137-145(1993).

CC -I- FUNCTION: Unknown.

CC -I- TISSUE SPECIFICITY: Limited to cell-lines of leukemic origin.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; M33509; AAA35585.1; -

CC EMBL; M33518; AAA35586.1; -

CC EMBL; M33512; AAA35586.1; JOINED.

CC EMBL; Z15025; CAA78744.1; -

CC FIR; B35098; B35098

CC Genew; HGNC:13918; BAT2.

CC MIM; 142580; -

CC Repeat.

CC KW DOMAIN 519 524 POLY-PRO.

CC FT DOMAIN 636 657 GLN-RICH.

CC FT DOMAIN 684 688 POLY-PRO.

CC FT DOMAIN 699 704 POLY-PRO.

CC FT DOMAIN 814 821 POLY-PRO.

CC FT DOMAIN 1340 1345 POLY-GLY.

CC FT DOMAIN 1398 1403 POLY-GLY.

CC FT DOMAIN 1436 1442 POLY-PRO.

CC FT DOMAIN 1982 1991 POLY-PRO.

CC FT DOMAIN 41 1795 4 X 57 AA TYPE A REPEATS.

CC FT REPEAT 41 95 1-1.

CC FT REPEAT 98 154 1-2.

CC FT REPEAT 281 337 1-3.

CC FT REPEAT 1740 1795 1-4.

RP RC
RC STRAIN=Berkeley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H.C., Blazek R.G., Nelson C.R., Miklos G.L.G.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Andrews-Pfannkoch C., Baldwin D.,
RA Abil J.F., Agbayani A., An H.-J., Bayraktaroglu L., Beasley E.M.,
RA Balow R.M., Basu A., Baxendale J., Berman B.P., Bhandari D., Bolshakov S.,
RA Beeson K.Y., Benos P.V., Berman B.P., Brokstein P., Brottier P.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo C., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K.J., Evans G.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Gloddek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harri N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mout S.M., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Nelson D.R., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Palazzolo M., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Reinart K., Remington K., Simpson M., Skupski M.P., Smith T.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Strong R., Sun E.,
RA Spier E., Spradling A.C., Stapleton M., Turner E., Wang A.H., Wang X.,
RA Swirskas R., Tector C., Turner R., Venter E., Weissbach J.,
RA Wang Z.-L., Wasarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
ET "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
[3]
RN FUNCTION, AND DEVELOPMENTAL STAGE.
RP MEDLINE=99112962; PubMed=9895321;
RX Vazquez M., Moore L., Kennison J.A.;
RA "The trithorax group gene osa encodes an ARID-domain protein that
RT genetically interacts with the brahma chromatin-remodeling factor to
RT regulate transcription.";
RL Development 126:733-742(1999).
[4]
RN DNA-BINDING, AND IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM AND SNR1.
RP MEDLINE=20069333; PubMed=10601025;
RX Collins R.T., Furukawa T., Tanese N., Treisman J.E.;
RA "Osa associates with the Brahma chromatin remodeling complex and
RT promotes the activation of some target genes.";
RL EMBO J. 18:7029-7040(1999).
[5]
RN FUNCTION.
RP MEDLINE=99403006; PubMed=10471712;
RX Staehling-Hampton K., Ciampa P.J., Brook A., Dyson N.;
RA "A genetic screen for modifiers of E2F in Drosophila melanogaster.";
RL Genetics 153:275-287(1999).
[6]
RN IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM; OSA; MOR; SNR1; DALAO;
RP BAP55; BAP60 AND BAP47, AND FUNCTION AS COACTIVATOR.
RX MEDLINE=20270023; PubMed=10809665;
RA Kal A.J., Mahmoudi T., Zak N.B., Verrijzer C.P.;
RT "The Drosophila brahma complex is an essential coactivator for the
RL trithorax group protein zeste.";
RN Genes Dev. 14:1058-1071(2000).
[7]
RN FUNCTION AS A COREPRESSOR.
RP MEDLINE=20573925; PubMed=11124806;
RX

QY 1973 SSPSKGSEPRPLPPVNSG-----HATIAITPAK-----NLAPHAS----- 2008
Db 1513 SSDPHFEPGPMVRGVGTFRDSAGVSPFPKRRPRPKPELLQESLPFPHSSGFLGS 1572
QY 2009 -PDPPAPAGASD-----PH-----REKTSQKPSIQ-----ELE 2037
Db 1573 KPEGFGQASRDGTGTEALPHIWNRLHTATSRKSYRPTSMPEWBEPLSPFEDVAGTEMS 1632
QY 2038 LRSLYGHSSVSPGVPSPVSPSITHKGL-----PKH-----LEELDKSHLEG 2084
Db 1633 QSDGVLDSGDSVSSGPCSORSCP-----DGLKGAAEGPKRPGSGSPNANVPCGGPG 1688
QY 2085 ELRQKQKPVKLGCEAAHLPHRLPSPSQSSPSSLLQTAQGVKHQVWVTLAQHISEVIT 2144
Db 1689 SEPPRRPPAPHDGDKELPREQLP-----FGPIGT-----SQR-----T 1726
QY 2145 QDTRHHPQQLSAPLAPLYFPGASCVPLDLR-----PPS-DL 2183
Db 1727 DRGTEPGPIRPS-HRPGPPVQF-GTSDKXDSLRLVWGDLSLKAKELTASVTEAIPVSRDW 1784
QY 2184 YLPPDPHAGAPAGSPHS-----EGGKSPSPN-----KTSVLGGGDEGI- 2222
Db 1785 ELLP-----SAAASAEPQSKNLDGCHCPSPSSGQRLPYEVYFGSAGSSQISGSHGLS 1841
QY 2223 -----EPVSPPEGMTBPGHRSRAV-YPLLYRDGEQTEPSPRMSGKS 2261
Db 1842 ITSQWRLRPGTSLHPYRSQPLXLPAGPAPPSALLSGVAKGQFLDFTMQATELGKLP 1901
QY 2262 PNTSOPPAFF-----SKTENSAMVSKQKEINKLTHNRNEPEYNISQPGTEI 2313
Db 1902 AGGVLYPPSPFLYSPAFCSPLDFTSLQVRQ-----DLFSP-SDF 1941
QY 2314 FNPALITG--TGLMTYRSQAVEHASTNGLAIIRKALMGKYDOWERSPLSANAFNPL 2371
Db 1942 YSTFLOQGGQSGFLSGFAQQ-----MLLPWDSQLPVNFGSLPPAPP---PAPPPL 1992
QY 2372 NASASLPAAMPITAADGRSDHTLTSFGGGKAKVGRSPSRKAKSPAGLSDGRRPSPVS 2431
Db 1993 SLLPVGALQPPSLA-----VRPPAPATRVLPSPFA-----RPPPAS 2029
QY 2432 SVHSE-----GDCNRRTPLTNRVWEDRPPSAGSTPFPYPLI--M 2469
Db 2030 LGRALHPVELKPFQDYQKLSSMLGPGSRTPPTGCRSFGSLNRLKATPSTYSYGVFTQ 2089
QY 2470 RLQAGVWASP-----PPFGLPAGCPLAGPHAWDEEP 2502
Db 2090 RVDLYQASPPDALRWPKPWERTGPPREGPSRR-AEEP 2128
RESULT 9
OSA DROME
ID OSA DROME STANDARD; PRT: 2716 AA.
AC Q81N94: O61603; Q9VEG7;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Trithorax group protein OSA (Eyeless protein).
GN OSA OR ELD OR CG7467.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX MEDLINE=97415319; PubMed=9271118;
RA Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
RT "eyelid antagonizes wingless signaling during Drosophila development
RL and has homology to the Bright family of DNA-binding proteins.";
RN Genes Dev. 11:1949-1962(1997).
[2]

RA Collins R.T., Treisman J.E.;
 RT "Osa-containing Brahma chromatin remodeling complexes are required
 for the repression of wingless target genes.";
 RT Genes Dev. 14:3140-3152(2000).
 RN [8].
 RP FUNCTION AS A COREPRESSOR, AND INTERACTION WITH PNR AND CHI.
 RX MEDLINE=22515897; PubMed=12629041;
 RA Heitzler P., Vanolst L., Biryukova I., Romain P.;
 RT "Enhancer-promoter communication mediated by Chip during
 Panier-driven proneural patterning is regulated by Osa";
 RL Genes Dev. 17:591-596(2003).
 CC -!- FUNCTION: Trithorax group (trxG) protein required for embryonic
 segmentation, development of the notum and wing margin, and
 photoreceptor differentiation. Required for the activation of
 genes such as Antp, Ubx and Eve. Binds to DNA without specific
 affinity, suggesting that it is recruited to promoters by
 promoter-specific proteins. Essential component of the Brahma
 complex, a multiprotein complex which is the equivalent of the
 yeast SWI/SNF complex and acts by remodelling the chromatin by
 catalyzing an ATP-dependent alteration in the structure of
 nucleosomal DNA. This complex can both serve as a transcriptional
 coactivator or corepressor, depending on the context. Acts as an
 essential coactivator for Zeste, which recruits the whole complex
 to specific genes. In contrast, it acts as a corepressor for Wg
 target genes, possibly via an interaction with Pan and Gro. It
 also acts as a negative regulator for proneural achaete-scute,
 when it is directly recruited by Pan and Chi. Also represses E2f
 activation.
 CC -!- SUBUNIT: Component of the Brahma complex, which is composed of
 Brm, Osa, Mor, Snr1/Bap45, Bap11/Dalao, Bap55, Bap60 and Bap47.
 CC -!- SUBCELLULAR LOCATION: Nuclear.
 CC -!- TISSUE SPECIFICITY: Ubiquitously expressed in early embryo. In
 third instar larvae, it is ubiquitously expressed in wing and eye-
 antenna imaginal disks, with a stronger expression in a band just
 anterior to the morphogenetic furrow.
 CC -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
 CC -!- DOMAIN: The ARID domains mediates the binding to DNA.
 CC -!- SIMILARITY: Contains 1 ARID domain.
 CC -!- SIMILARITY: Contains 1 EHD (Eyelid homology) domain.
 CC -!- CAUTION: Ref.2 (AA055457) sequence differs from that shown due to
 erroneous gene model prediction.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF053091; AAC06254.1; --
 DR EMBL; AE003718; AAF55457.1; ALT_SEQ.
 DR EMBL; AE003718; AAN13750.1; --
 DR PIR; T13049; T13049.
 DR FlyBase; FBgn0003013; osa.
 DR GO; GO:0005634; C:nucleus; IDA.
 DR GO; GO:0003677; F:DNA binding; IDA.
 DR GO; GO:0046530; P:photoreceptor cell differentiation; IMP.
 DR GO; GO:0045449; P:regulation of transcription; IDA.
 DR GO; GO:0007379; P:segment specification; IMP.
 DR GO; GO:0008587; P:wing margin morphogenesis; IMP.
 DR GO; GO:0016055; P:Wnt receptor signaling pathway; IMP.
 DR InterPro; IPR001606; ARID.
 DR InterPro; IPR008938; ARM.
 DR InterPro; IPR006031; XYPXX.
 DR Pfam; PF01388; ARID; 1.
 DR Pfam; PF02162; XYPXX; 8.
 DR SMART; SM00501; BRIGHT; 1.
 DR Transcription regulation; DNA-binding; Activator; Repressor;
 KW Chromatin regulator; Nuclear protein; Developmental protein.
 FT DOMAIN 997 1111
 FT DOMAIN 1769 2517
 FT EHD.

FT	DOMAIN	19	1763	PRO-RICH.
FT	DOMAIN	174	380	GLN-RICH.
FT	DOMAIN	619	873	GLY-RICH.
FT	DOMAIN	1222	1453	GLY-RICH.
FT	DOMAIN	1271	1751	GLN-RICH.
FT	DOMAIN	1730	1745	HIS-RICH.
FT	DOMAIN	2589	2624	SER-RICH.
FT	DOMAIN	2625	2716	ALA-RICH.
FT	CONFLICT	61	61	MISSING (IN REF. 1).
FT	CONFLICT	1169	1169	V -> G (IN REF. 1).
FT	CONFLICT	1795	1795	M -> T (IN REF. 1).
FT	CONFLICT	2637	2637	G -> E (IN REF. 1).
SQ	SEQUENCE	2716 AA;	284063 MW;	EFAE76CB51C7C675 CRC64;

Query Match 3.8%; Score 503; DB 1; Length 2716;
 Best Local Similarity 20.6%; Pred. No. 3.3e-09;
 Matches 440; Conservative 195; Mismatches 776; Indels 724; Gaps 108;

Qy	670	KLKMEKERNARRKKKKAPAA-----ASEEAAFPVVEDEEMEAASGVSGNBEEMVEEABA	723
Db	4	KIKSPQTOQQOQCGAPAPAAATPPSAGAAPGAATPPT-----	39
Qy	724	LHASGNEVPRGECGPGATVNNSSDTESIPS-----PHTAAKDTGQNGPKPPATLG	774
Db	40	-----SGPPTNNNNNGSDPSIQOQQQNVAPHYGA-----PPPPGS--	77
Qy	775	ADGPPPGP-----PTPRRTSR-----APIETPASEATGAP-----	806
Db	78	PGGPGPGDPDAVMHYHLHQOQQOHHPPPHMQOQHHGGPAPPPPGAPGAPHAPOVKKEY	137
Qy	807	--TTPPPAP-----PSPSAPPVVPVVEEETAAAPVVEE--GE	841
Db	138	THLPHPHPHYPAYHADPNMDPVRYGQLPGGKPP--QQQPHPPQOQPPQPPGPGGS	193
Qy	842	EOKPAPAEELAVDTGKAEPEVKSECTEAEERP-----AKGDAEAEAETAGLAKAEK	895
Db	194	PNRPPQOQRYIPGQPPQGTPTLNSLSSNPPPPQHRYANTYDPQAAAAAQAQQ	253
Qy	896	KEGSGRATTAKSSGAPQSDSSATCSADEVEAEGGDKN-----RLLSR-----	941
Db	254	QOAG-GPPPHGHPPPHQPSPS-----PYGQOQGWAPPPIPSYQOLGSPSOYR	301
Qy	942	--PSLLTPTGD--PRAN-----ASPKPLDLKQLKRAAAI-----PPIQVT	979
Db	302	TPPTNTSRGQSPYPAHGQNSGVSYPSSQQOQQOQQOQQOQPGVPGVPGPPGTG	361
Qy	980	KVHEPPREDAAATKAPAPAP--PPQNLI-----QPESDAFQQPGSSSPRG	1021
Db	362	Q--QPQOQNTPTTSQYSFYQRYPTPPLPAGGSGNHRHTAYSTHQYPEPNRPWPGSSSP	419
Qy	1022	KSRSPAPADKEAFABAQKLPGDPPCWTGCLPFPVPPREVIKASP-HAPDPSAF-SYAP	1079
Db	420	GSGHPLPPASPH-HVPLPQQOQPPPPPHVSAGP--PP-----SSSPGHAPSPSPSOAS	471
Qy	1080	PGHPLPL-----GLHD--TARVLPFRPTTISNPPLISSAKHPSVLERIGAI	1125
Db	472	PSPHQELIGQNSNDSSGGAHSGMGSGPGCTFNQQVMRPTTSPSTGSS-----GSR	522
Qy	1126	SQMSVOLHVPYSEHAK-----APV--GPVTMGLPLPMDPKKLAPSGVQESLSPRGQ	1177
Db	523	SMSPAVAQNHPISRPASNSQSSGGGPMQPPVYAGAGGPPMPHPMPMP-GGPPQQOQSQOQ	581
Qy	1178	AGPPESLGVPTAOEASVLRGTALGSLVPGGSIITKGIPTSTRVPSDSAITVRGSIITHTPADV	1237
Db	582	ASNSASSASNSPQOT-----PPAPPNNQNMNNMATPPPPP--QGAAGGYPMP	629
Qy	1238	LYKGTITRIIGEDSPSLDRGREDSLPKGHVIEGKKGHLVLSYEGGMSVTCQSKEDGRSS	1297
Db	630	HHGGYK--MGGPGQSPGAQGYPPQQPQ-----QYPPGNVYPPPPYPPPGAYA	674
Qy	1298	SGPHETAAKRTTYDMMEGRVRGIRAISSASIEGLMGRAIPPE--RHSPHLL--KEQHHR	1352
Db	675	TGPP-----PPPTSQAGAGGANSMPGSAQAGGYPGRGMPNHTGTQYPPYQWVPPSPQQTVP	729

1353 GSITQIGPRSYEAQEDYLREAKLLKREGTTPPPPPSRLDTLAYKTQALGPLKLKPAHE 1412
 730 G---GAPGGAMVGNH-----VQKGTTPPPVVG-----GP-----PPQ 760
 1413 GLVATVEAGRSIHIEPRELHTPELPLAPRLKEGSIQTGTPKLYDTGASTGSKKD 1472
 761 G-----SGSPRLNLYL-KOHLQHGKGGYSGPTP-----PQG-PQGVNG--PTGM----- 801
 1473 VRLSIGSGRTFPVPH-----PLDVMADARALERACRYEELSKSRRECTASSG 1519
 802 -----HPCMPGPHMGPHGPTNMGPPTSTPQSQMLQ-----GGQPGQGSAG 847
 1520 GSIARGAPVIVPELGRKQPSLTVEDHGAPAGHLPRGSPVTMREPTPRLOEG--SLSSS 1577
 848 G-----PESGGPEH-----ISQDNGISSG--PTGA-AGMHAVTSVVTTGPDGTSMD 891
 1578 KASQDRKLTSTPREIAKSPHSTVPEHHPHPISPYEHLLRGVSGVDLYRSHIPLAFDPTS1 1637
 892 EVSQOSTLSNAAASGEDPOCTTPKSKN-----DP----- 922
 1638 PRGIPDLAAAYLPHRLAPNPTYPH---LYP-----PYLIRGYPDPTAALENRQTIIND 1688
 923 -----YSQSHLAPSTSPHVVMPHGGGPGGEYDMSSPPNPNRAGSPQVFENH 970
 1689 YITSQMHNTATAMAQADMLRGLSPRESSLALNAAAGRGID-----LSQVHLPLV 1744
 971 VVPQEPFRSTITT-TKSDSLCKLYEMDN-----PDREGWLDKLRAMEERETPLTA 1023
 1745 PTPGTPTAMDRLAYLTPAQPPSSRSHSSPLSPGPTHTLTKPT-----TSS 1793
 1024 CPTISKQPLDLRL-YIYVYKER-----GGFVETKSKTKDIAGLLIGASS 1069
 1794 SERERDRDRDRDRERKSILT-----STTTVEHAPIWR---POTEGSSGSSG--SGG 1843
 1070 SAAITLR-----KHYYKNLITFECHDRGIDPLPIIQVVEAGSKKTKAKAASVPSPG 1122
 1844 GGGSSSRPASHGA-HQHSPIGPTQDALQORPSVLHNTGMK-----LIITAVE 1891
 1123 GG-----HLDAGTTNSTGSSNQDSFPAPPGSAPNAADIGYFGYGGGSPYVVASGQ 1174
 1892 BSKPTVLRSTSSVVRPAATPPATHCPLGCTLDGVYPTLMPEVLLPKPEAPRVARPERP 1951
 1175 PDVATAGQORPPSQNNPOTPHGAAVAAG--DNI-----SVSNPFEDP-IAAGGP 1225
 1952 RADTGHAFILAKPPARSGLEPASPP-----SKGSEPRPLVPVSGHATIAITPA---KN 2001
 1226 GSCTG-----PGQGGPGGAASGAGAVGAGVGPGQPHPPPHSPHTAAQQAAGQHQ 1279
 2002 LAPHASPPDPPAPPASADPHREKTSKPFISQLELSLGLVHGSYSPE---GVEPVS 2057
 1280 QHPQHQHGLGFPFPQQQGGGQGGQPPP-----SVG--GGPPAPQQRHGPQVPPS 1329
 2058 PVSSPSLTHDKGLPKHLELDKSHLEGELRPQGPVKGGEAAHLPHLRPLPESQPSSS 2117
 1330 P-----QHVRAAGAVYPPGSG-----YPTFVSRTPGS 1359
 2118 PLLOTPAGVKGHORVTLAQHISEVITQDTRHHPOOLSAPLPAPLYSPFGASCPVLDLR 2177
 1360 P-YPSQPGAYG-----QYGSDDQYNATGPPQPGQPGPG-----Q 1393
 2178 RPPSDLYL-PP--PDHGAPARGS-----PHSEGGKRSRPPNKTSLVGGGEDIEBV 2225
 1394 YPQNRNMYPPYGEAGEAPTGANQVGYGSRYSQPPPGPQPPPTQTVAGGPAGGAPG 1453
 2226 SPPEGMTEPGHRSASVYLLYRDEQTEPRSMGSKPGNTSQPPAFPSKLTESNANVKS 2285
 1454 APP-----SSAYP-----TGRPSQDDYQPPPPQSPQPRRHPDFIKD 1490
 2286 KQBEINKLNTNNEPEYNI SQGTBEI FNWPAITG---TGLMTYSQAVQEHASTMGL 2342
 1491 -----SQPYPGYNA-----RPQIYGAWQSGTQYRPQYPSAPQN--- 1526

2343 EAIIRKALMGKYDWEESPPLSANAFLNASALPAAMPITAAAGRSDDTLTSPGGGK 2402
 1527 -----WGGAPP--RGAAPPAGHPPIQOP--AGVAQWDQHRYPPOQGP 1568
 2403 -----AKVSGRPSRKAKSPA-----PGLASGD-----RPS 2429
 1569 PFPQOQOQPPQOQOQPPYQVAGPQPPQAPQOWAQNPGQTAAOSGIAPPGSLRPPS 1628
 2430 VSVHSEGDGCRERTLTNRVWEDRPSASGTFPP-----YNPLIMRLQAGVM--- 2476
 1629 -----GPGQONRNPGPAQ--OQSQOQGGVPQPPQASHGVSFGLPQVPGMGVMPK 1681
 2477 --ASPPPP-----GLPAGSGLAGPHHAWDEEPKPL 2505
 1682 PYAMPPPPSGQVGVQVQGP-----PGGMMSQKDPFM 1713

RESULT 10
 MAPA RAT
 ID MAPA RAT STANDARD; PRT; 2774 AA.
 AC P34926;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Microtubule-associated protein 1A (MAP 1A) [Contains: MAP1 light chain LC2]
 GN MAP1A.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain.
 RX MEDLINE=92355629; PubMed=1379599;
 RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
 RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in one messenger RNA.";
 RL J. Biol. Chem. 267:16561-16566(1992).
 CC -!- FUNCTION: Structural protein involved in the filamentous cross-bridging between microtubules and other skeletal elements.
 CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate with MAP1A and MAP1B proteins.
 CC -!- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE THEIR MORPHOLOGY.
 CC -!- DOMAIN: The basic region containing the repeats may be responsible for the binding of MAP1A to microtubules.
 CC -!- PTM: Various serine residues may be phosphorylated by cAMP kinase.
 CC -!- PTM: LC2 IS COEXPRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH BOTH MAP1A AND MAP1B.
 CC -!- SIMILARITY: TO MAP1B.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.ebi.ac.uk/ebis-sib.ch/announcement/> or send an email to [license@sib-sib.ch](mailto:license@sib-sb.ch))
 CC EMBL; M83196; A848069.1; -;
 CC PIR; A43359; A43359.
 DR Microtubule; Repeat; Phosphorylation.
 KW CHAIN ?2465 2774 MAP1 LIGHT CHAIN LC2.
 FT REPEAT 11 X 3 AA REPEATS OF K-K-[DE].
 FT DOMAIN 309 496
 FT DOMAIN 336 541
 FT REPEAT 336 338
 FT REPEAT 415 417
 FT REPEAT 420 422
 FT REPEAT 424 426


```
Db 2108 WPETAYSSLSHSLGVRSLDFPASAFSSLOPAP--PQL--PSPAPRSPAPCGSL 2163
Qy 1756 ---DR-LAYLPTAPQPFSSRHSS-----SPLSPGGP--THLTXTPT 1790
Db 2164 AFSGRALALVPGTPT--RTHDEYLVKTKAPSLDSSLQPLPSPSPGGPLSLNLRPAS 2221
Qy 1791 TSSSERDRDRDRDREREKSLTSTTTVEHAPIWRPTEQSSGSSGGGGGSSSR 1850
Db 2222 FALSE-----GSSSEATTPISSVAERFP--PGLE-----AAEQSAEGLGSGKE 2263
Qy 1851 PASHSIAHQHPISPRITODALQOPSVLNTGKMGITAVEPSKPTVLRTSTSSPVRPA 1910
Db 2264 SAAHS-LWDLTFLSPASASDLAPA-----PA 2290
Qy 1911 ATFFPATHCPGLGTLGDVYPTLME-----PVLLPK-----EAPR-----V 1945
Db 2291 PAPAPAGLP-GDLGDGTLPCRPCTGELTKKPFLLSPSGDHEANGPGETSLNPPGFVT 2349
Qy 1946 ARPERPRADTGHAF-----LAKPPARSGLPEPASSKSGSEPRPLVPVSGHATI 1994
Db 2350 ATAEEEAEPHAWERGSGWPEGAERSSRPDTLLSSEQPLRPGKSGG-----PPCSLSSEV 2405
Qy 1995 ARTP---AKNLAPH--HASP---DPPAPPASADPHREKTSQKPSIQELLE-----RS 2040
Db 2406 EAGPGCATDPRPHCGELSPFLNPLPFS-----TDDSLSTEARELAGKGGRR 2456
Qy 2041 LQYHGSSYSP-----EGVFPVSPVSPSLTHDKPLHLELDKSHLEGLRKPQGP---P 2093
Db 2457 VGRPATGCPMADETPTTSASDSSGSDSDVPPEECPSITAEALDSDSDGDFLP 2516
Qy 2094 V-KLGG-----EAAHLPHLRPLPESOPSSPLLQTPAGVKGHQRVTLAQHISEVITQ 2145
Db 2517 VDKAGVSGTHPRPGCHDPPPTPLDPRPSP-----RP 2550
Qy 2146 DYTRHHPPQOLSAPLAPLYSPFGACPVLDLRRPSPDLPLPPDHGAPARGSPHSEG--- 2202
Db 2551 DVMCMADPEGLS-----ESGRVERLEKCR-----PGERAPGRAPKSPARRL 2593
Qy 2203 ---GKRSPFPNKTSLVGGEDIEFVSPPEGMT-----BFGH---SRSAVYPLLVRDGBQT 2252
Db 2594 DIRGKSPFTFGKVPDRTSRTVPRSTPQVTSAAEKDGHSPMSKGLV-----NGLKA 2647
Qy 2253 EFSRMGSKSPGNTSQPPAPFASKLTESNANWKSKEINKLNTH-----NRNEP 2302
Db 2648 GSTALGSK---GGSGPPVYDLYATPNHCSGKTADQDFRRVRASYVYVVGNDP 2698

RESULT 11
MML2_HUMAN
ID MML2_HUMAN STANDARD; PRT; 5262 AA.
AC O14686; O14687;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Myeloid/lymphoid or mixed-lineage leukemia protein 2 (ALL1-related protein).
GN MML2 OR ALL1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RX MEDLINE=97388474; PubMed=9247308;
RA Prasad R.; Zhadanov A.B.; Sedkov Y.; Bullrich F.; Druck T.;
RA Rallapalli R.; Yano T.; Alder H.; Croce C.M.; Huebner K.; Mazo A.;
RA Canaan E.;
RT "Structure and expression pattern of human ALL1, a novel gene with
RT strong homology to ALL1-1 involved in acute leukemia and to Drosophila
RT trithorax."
RL Oncogene 15:549-560(1997).
RN [2]
```

```
RP INTERACTION WITH ASC-2/NCOA6 CONTAINING COMPLEX.
RC TISSUE=Cervical carcinoma;
RX MEDLINE=22371436; PubMed=12482968;
RA Goo Y.-H.; Sohn Y.C.; Kim D.-H.; Kim S.-W.; Kang M.-J.; Jung D.-J.;
RA Kwak E.; Barlev N.A.; Berger S.L.; Chow V.T.; Roeder R.G.;
RA Azores D.O.; Meltzer P.S.; Suh P.-G.; Song E.J.; Lee K.-J.; Lee Y.C.;
RA Lee J.W.;
RT "Activating signal cointegrator 2 belongs to a novel steady-state
RT complex that contains a subset of trithorax group proteins.";
RL Mol. Cell. Biol. 23:140-149(2003).
CC -1- FUNCTION: May be involved in transcriptional regulation.
CC -1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which
CC contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/
CC RBBP5, alpha- and beta-tubulins, the trithorax group proteins
CC MLL2 and MLL3, and ASH2/ASCL2.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Name=1;
CC IsoId=O14686-1; Sequence=Displayed;
CC Name=2;
CC IsoId=O14686-2; Sequence=VSP_008563, VSP_008559;
CC Name=3;
CC IsoId=O14686-3; Sequence=VSP_008560;
CC -1- TISSUE SPECIFICITY: Expressed in most adult tissues, including a
CC variety of hematopoietic cells, with the exception of the liver.
CC -1- MISCELLANEOUS: This gene mapped to a chromosomal region involved
CC in duplications and translocations associated with cancer.
CC -1- SIMILARITY: Belongs to 5 PHD-type zinc fingers.
CC -1- SIMILARITY: Contains 5 PHD-type zinc finger.
CC -1- SIMILARITY: Contains 1 post-SET domain.
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.
CC -1- SIMILARITY: Contains 1 SET domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; AF010403; AAC51734.1; -.
CC EMBL; AF010404; AAC51735.1; -.
CC PIR; T03454; T03454.
CC PIR; T03455; T03455.
CC Genew; HGNC:7133; MLL2.
CC MIM; 602113; -.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0003700; F:transcription factor activity; TAS.
CC GO; GO:0007048; P:oncogenesis; TAS.
CC GO; GO:0006366; P:transcription from Pol II promoter; TAS.
CC InterPro; IPR003889; FYrich_C.
CC InterPro; IPR003888; FYrich_N.
CC InterPro; IPR000910; HMG_12_box.
CC InterPro; IPR003616; PostSET.
CC InterPro; IPR006118; Recombinase.
CC InterPro; IPR001214; SET.
CC InterPro; IPR001965; Znf_PHD.
CC InterPro; IPR001841; Znf_ring.
CC Pfam; PF00628; PHD; 5.
CC Pfam; PF00856; SET; 1.
CC SMART; SM00542; FYRC; 1.
CC SMART; SM00541; FYRN; 1.
CC SMART; SM00398; HMG; 1.
CC SMART; SM00249; PHD; 7.
CC SMART; SM00508; PostSET; 1.
CC SMART; SM00184; RING; 3.
CC SMART; SM00317; SET; 1.
CC PROSITE; PS50868; POST_SET; 1.
CC PROSITE; PS50280; SET; 1.
CC PROSITE; PS01359; ZF_PHD_1; 5.
CC PROSITE; PS50016; ZF_PHD_2; 5.
```


Db 2110 PRQPPPPSCCLPPRS-LPSDPFSRVVPSQSSQSSQLTPRPLSAEAF- 2161
Qy 1044 GDPCWTSGLPVPVPPREVITKASHPADPSAFSAVPPGHL-PLG-LHDTARVLP RP 1099
Db 2162 -PSPVTPRF-----QSPDPYS--RPPSRPQSRDPFPAPLH--KPPRPQP 2199
Qy 1100 PTISNPPLTSSAKHPSVLEROIGAI SQGMSVQLHV-PYSE-HAKAPVG-----PVT 1149
Db 2200 PEV-----AFKAGSLAHTSLGA--GGFPAALPAGAGELHAKVPSGPPNFVRSPGT 2249
Qy 1150 -MGLPLPM-----DPKLAFTSGVKQEQLSP-----RGOAGPESIGVPTAQEAS 1193
Db 2250 GAFVGTSPMRFTFPQAVGESLXPP--VPQGLPPHGINSHFPGPTLGKPEQSTNYT 2306
Qy 1194 VL-----RGVALGSVPFGS--ITKGIPISTRVPS-----DSAITYRGSIHTHGPADVLY 1239
Db 2307 VATGNFHPGSPGLFGSSGTGESYGLSLRPPSVLPPAPDGLPY--LSHGASQ--R 2360
Qy 1240 KGTITRIIGEDSPRLDRGREDSLPKGHVIVGKKGHVLSYEGGMSVTOCKED----- 1293
Db 2361 SGITSPVEKREDPG--TGMGSSLAETAE--PGTQDPGMS--GLSOTELKQKQORLRL 2412
Qy 1294 -GRSSGPPHETAAKPYDMMEGRVGRATSSASIE-----G 1329
Db 2413 ELLTROQIORTLROEKETAAAGACVPPGSGWGAEPSSPAFEQLSRGOTFFACTQDKSS 2472
Qy 1330 LMGRAPIPRHSPHLLKEQHIRGSIOTGIPRSVVEAQEDVLRREAKLLKREGTPPPP-- 1387
Db 2473 LVG-LPPSK-----LSGPILG--PGSF--PSDDLRSR-----PPPPAT 2505
Qy 1388 PPSRDL-----TEAVKTOA--LGPLKKAHEGL-----VATVKEAGRSIH--EI 1428
Db 2506 PSSMDVNSRQVGGSQAPYQYAPYVPSLPLQOQQOQLWQOQOATAATSMRFAMARFPST 2565
Qy 1429 PREEL-RHTPELELA-----PRPLKEGSIQTGPTLKYDTGASTTGSKKHDVRLSIGSPG 1481
Db 2566 PGPELGRQALGSLAGISTRLPGP-----GEPVPGPAGPAQFIELRHNVQKGLGPGG 2617
Qy 1482 RTFPVPHLDMADARALERACYEBSLKSRTGATSSGSGSTARGAPVIVPELGRQSPPL 1541
Db 2618 TFPF-----TEAVKTOA--LGPLKKAHEGL-----VATVKEAGRSIH--EI 1428
Qy 1542 TYEDHGAFFAGHLPRGSPVTMRPTPLQEGSLSSKASQDRKLTSTPREIAKSPHSTVP 1601
Db 2635 SEDPH-----RLAPEGU-----GLAVSGLPPQKPSAPPAP-ELNLSLHPT-- 2674
Qy 1602 EHHPHPISP-----YEHLLRGVSGVDLYRSHIPLAFDPTSIPIRGIP-----LDAAAAYYL 1651
Db 2675 ---PHTKGTPLTGLELVNRPPSSTELGRPN-PLALEAGKLPCEDPELDDDDFDAKHALED 2730
Qy 1652 PRHLAPNTPYHL-YPPVYLIRGYPTAALENRQT----- 1684
Db 2731 DEELA-----HLGLGVDAVAKGDBELGTLENLTNDPHLDDLNGDEFLLAYTDPDLDT 2784
Qy 1685 -----IINDYITSQQMHNTATAMAQRAADMLRGLSPRESSLALNVAAGPRGIIDLSQVPH 1739
Db 2785 GKDKDIFNEHLVBE-----SANEAEERALLGVPE-----GFLG----- 2820
Qy 1740 LPVLVPPPTPGTATAMRLAVLPTAPQPF--SSRHSSSPLSFGPGTHLTKTPTTSSSRE 1797
Db 2821 -PEERPP-PAADASEPRLASVLPEVKPVKEGGRHPS-----PCQFT----- 2860
Qy 1798 RDRDRDRDREREKSIITSTTVEHAPI-----WRPG-----TEQSSGSGSGG 1842
Db 2861 -----IATPKVPEPAANSLGLKPKGQSMMSGRDTRMGTPGFSSG 2902
Qy 1843 -----GGGSSSRPASHSHAHQHSPISPRTQDALQORPSVLHNTGMKGIITAVEPSKPTV 1897
Db 2903 HTAEKASGATCGPPAH-----LLTSPSLSGPGSSSLEKFE----- 2939
Qy 1898 LRSTSTSSVRPAATPPPPATHCPLGGLTLDGVVPTLMPEVLLPKPEAPRVARP--ERPRADT 1955
Db 2940 LESGALTLPGGPAAS-----GDELD-----KMESSLVASELPLLTIDLEHEKKE- 2984

Qy 1956 GHAFIAKPPARSCLGEPASSPSKSGSEPRPLVPPVSGHATIAARTPAKNLPHHAGSPDPAPP 2015
Db 2985 ---LQKKQQLSALQQLPAQOQQOQQOQHSLLP-----AEGPA 3017
Qy 2016 ASASDPRHREKTSQKPFISIQLERLSLGHSSYSPEGVEPVSPVSPSLTHDKGL-PKHL 2074
Db 3018 QAMSLPHEGSSPSLAGSOOQL---SLGL-AVARQPGLPQLPMTQPPAHALQORLAPSMA 3073
Qy 2075 EELDQKH-LEGEURPKQPGVVKLGGEANHLPHLRPLPESOP--SSSPLLQTAQVKGHOR 2131
Db 3074 MVSNQGHMLSGQ-----HGGQAG---LVPOQSSQPVLSQKPMGTMPMSMCKPQ 3119
Qy 2132 VVTLAQHISEVI--TQDYTRHHPOQLSAPL-PAPLYSFPGAS-----CPVLDLRR 2178
Db 3120 QLAQQOQLANSFPDPTDLKFAEDIIIGPIAKAKVWALKIKKVMAGSGISGVAFGMRQQ 3179
Qy 2179 ---PPSDLYLPPPDHGAFGARPSHSGEGRKSRPEPNKTSVLGGGEGDIEVPSP 2227
Db 3180 VSLLAQRLSGPSSDL---QNHVAGSGQERSAGDPSQPRPNPPTFAQGVINEADQRY 3235
Qy 2228 PEGMTEPHGSRSAVYPLL-----YRGEQTEPSRMGSKSPGNTSQPPAPFSKLTES 2278
Db 3236 EEWLF---HTQQLIQMLQVLEEQIGVHRKSRKALCAKQRTAKKAGREFPEADAELK--- 3289
Qy 2279 NSAMVKSQKOBINKLNTNRNEPEYNI SQPTEIFNMPAITGTGLMT-YRQAVQSEHAS 2337
Db 3290 --KLVTQOQSKIQOLQOVVKQKEH-----TNLMAEYRNKQOQQO-- 3328
Qy 2338 TNMGLEAIIRKALMGKYDQWEEPSPLSANAFNPUNASASIPAAIMPITAAAGRSHTTSP 2397
Db 3329 -----QOQQOQQOQHSAVLALSP--SQSPRLTLKLPGLLPG--HGLQPP 3369
Qy 2398 GGGKAKVGRPSRKAQSPAGLASGDRPPSVSVHSEGDGNCNRTPLTNRVWEDRPSA 2457
Db 3370 QG-----PPGQAGG----- 3379
Qy 2458 GSTPFPYPLIMRLQAGVMASSPPPLGLP-----AGSGPLAGP 2494
Db 3380 -----LRLTPGMALPGPGGFFLNTALAQOQQOQHSAGSLAGP 3420

RESULT 12

SPEN DROME STANDARD; PRT: 5560 AA.
ID Q8SX83; Q9NH17; Q9U6C3; Q9VPL1; Q9VPL2;
AC 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Split ends protein.
GN SPEN OR CG18497.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4), FUNCTION, ALTERNATIVE PROMOTER
RP USAGE, TISSUE SPECIFICITY, DEVELOPMENTAL STAGE, AND MUTANTS E9 AND
RP D57.
RC TISSUE=Embryo;
RA MEDLINE=20025936; PubMed=10556062;
RA Wiellerte E.L., Harding K.W., Mace K.A., Ronshaugen M.R., Wang F.Y.,
RA McGinnis W.;
RT "spen encodes an RNP motif protein that interacts with Hox pathways
RT to repress the development of head-like sclerites in the Drosophila
RT trunk."
RL Development 126:5373-5385(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Embryo;
RX MEDLINE=20157049; PubMed=10655223;
RA Rebay I., Chen F., Hsiao F., Kolodziej P.A., Kuang B.H., Lavery T.,

RA Suh C., Voas M., Williams A., Rubin G.M.;
RT "A genetic screen for novel components of the Ras/mitogen-activated
RT protein kinase signaling pathway that interact with the yan gene of
RT Drosophila identifies split ends, a new RNA recognition motif-
RT containing protein.";
RL Genetics 154:695-712(2000).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1), SUBCELLULAR LOCATION, AND TISSUE
SPECIFICITY.
RC TISSUE=Embryo;
RX MEDLINE=20171275; PubMed=10704397;
RA Kuang B.H., Wu S.C.-Y., Shin Y.-A., Luo L., Kolodziej P.A.;
RT "split ends encodes large nuclear proteins that regulate neuronal
RT cell fate and axon extension in the Drosophila embryo.";
RL Development 127:1517-1529(2000).
RN [4]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=Berkley;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.P.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke S., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D.A., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster";
RL Science 287:2185-2195(2000).
RN [5]
RP REVISIONS, AND ALTERNATIVE SPLICING.
RX MEDLINE=22426069; PubMed=12537572;
RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradscky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN [6]
RP SEQUENCE OF 424-2002 FROM N.A.
RC STRAIN=Berkley; TISSUE=Embryo;

RX MEDLINE=22426066; PubMed=12537569;
RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA George R.A., Guarni H., Krommiller B., Pacleb J.M., Park S., Wan K.H.,
RA Rubin G.M., Celniker S.E.;
RT "A Drosophila full-length cDNA resource.";
RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN [7]
RP FUNCTION.
RX MEDLINE=20253107; PubMed=10790398;
RA Lane M.E., Blend M., Heidmann D., Herr A., Marzodko S., Herzig A.,
RA Lehner C.F.;
RT "A screen for modifiers of cyclin E function in Drosophila
RT melanogaster identifies Cdk2 mutations, revealing the insignificance
RT of putative phosphorylation sites in Cdk2.";
RL Genetics 155:233-244(2000).
RN [8]
RP FUNCTION ON EGF RECEPTOR PATHWAY.
RX MEDLINE=20414403; PubMed=10959845;
RA Chen P., Rebay I.;
RT "split ends, a new component of the Drosophila EGF receptor pathway,
RT regulates development of midline glial cells.";
RL Curr. Biol. 10:943-946(2000).
RN [9]
RP FUNCTION ON Wg PATHWAY.
RX MEDLINE=22668876; PubMed=12783785;
RA Lin H.V., Doroteo D.B., Cho S., Chen F., Rebay I., Cadigan K.M.;
RT "Split ends is a tissue/promoter specific regulator of Wingless
RT signaling.";
RL Development 130:3125-3135(2003).
RN [10]
RP FUNCTION: Probable corepressor protein, which regulates different
RP key pathways such as the EGF receptor and Wg pathways. Involved in
RP neuronal cell fate, survival and axon guidance, cell cycle
RP regulation and repression of head identity in the embryonic trunk.
RP May act with the Hox gene Deformed and the EGF receptor signaling
RP pathway. Positive regulator of the Wg pathway in larval tissues
RP but not in embryonic tissues. May act as a transcriptional
RP corepressor protein, which repress transcription via the
RP recruitment of large complexes containing histone deacetylase
RP proteins.
RN [11]
RP SUBCELLULAR LOCATION: Nuclear.
RN [12]
RP ALTERNATIVE PRODUCTS:
RX Event=Alternative promoter;
RX Comment=2 isoforms, 1 (shown here) and 2, are produced by use of
RX alternative promoters;
RX Event=Alternative splicing; Named isoforms=4;
RN Name=1;
RX IsoId=Q8SX83-1; Sequence=Displayed;
RN Name=2;
RX IsoId=Q8SX83-2; Sequence=VSP_008565, VSP_008566;
RX Note=No experimental confirmation available;
RN Name=3; Synonyms=SpentL;
RX IsoId=Q8SX83-3; Sequence=VSP_008567;
RX Note=Produced by alternative splicing of isoform 1;
RN Name=4; Synonyms=Spens;
RX IsoId=Q8SX83-4; Sequence=VSP_008565, VSP_008566, VSP_008567;
RX Note=Produced by alternative splicing of isoform 2;
RN [13]
RP TISSUE SPECIFICITY: Ubiquitous. Expressed prior to cellularization
RP in stage 3 embryos, and in blastoderm cells, including pole cells.
RP Expressed throughout the rest of embryogenesis. Later, it is
RP expressed at higher level in epidermal cells and CNS.
RN [14]
RP DEVELOPMENTAL STAGE: Isoform 3 is expressed both maternally and
RP zygotically.
RN [15]
RP SIMILARITY: Belongs to the Spn family.
RN [16]
RP SIMILARITY: Contains 3 RNA recognition motif (RRM) domains.
RN [17]
RP SIMILARITY: Contains 1 SPOC domain.
RN [18]
RP CAUTION: It is uncertain whether Met-1 or Met-7 is the initiator.
RN [19]
RP This SWISS-PROT entry is copyright. It is produced through a collaboration
RP between the Swiss Institute of Bioinformatics and the EMBL outstation -
RP the European Bioinformatics Institute. There are no restrictions on its
RP use by non-profit institutions as long as its content is in no way
RP modified and this statement is not removed. Usage by and for commercial
RP entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

Qy	1326	SIEGLMGRATPPERHS	PHLUKEQHIIHIGSITQ	IGIPRSYV-----	1366							
Db	3879	-----	QHNOCHLNQOLHAQHPT	KQHQAOQFNQOIQOHQSOQHQQVQQN	3926							
Qy	1365	EAQEDYLRREAKULL	KREGTTPPPPPS	RDLEAYKTOAL--GPLKLP	PAHEGLVATVKEAGR	1423						
Db	3927	QAQOQHUSQOQHOSQ	QQLNQOHQAQOOLQOI	QKLOQMHEG	QQQOKSPQG-----	VGHJGG	3982					
Qy	1424	SIHEIPREELRHTEPL	EL--	APRPLKEGSI	TQGTPLKYDYTCASIT--	-----	1466					
Db	3983	STSFASQO--	HNSQLPARGV	POQHPQOLSHSS	PCPKNTLVS	VNQGVPQPAILTRVGS	4040					
Qy	1467	-----	GSKGHDVRS	LIGSGRTPPPV	PHLDVMADARAL--	ERACVEESLKSR	1511					
Db	4041	SQPNQOQLPHQOSS	SGHPHQQLSSPGANL	PLQTPLNV	INTP	KTIIVQHIAQNVPP	4100					
Qy	1512	PGTASSGGSI	ARGAVIPE--	LGKPRQSP	LTYEDHGAP	FAGHLP	PRGSPVTMEPTPRLQ	1570				
Db	4101	PQT-----	QQNAIHYQ	QONQKOSTPP-----	GHV-----	EPTP----	4128					
Qy	1571	EGSLSSKASQDRKLT	STPREI-----	AKSPHSTV	PEHHPIIS--	PYEHLLRGVS--	1619					
Db	4129	--	AMSAQKTS	ESVSIVRPTTT	GLAVISANTV	GSLLTEENLIKISQ	KODELIEQDSKE	4186				
Qy	1620	-----	GVD--	LYRSHI	PLAFDPTSI	PGIPLDAAAYVL	PHLAPNTYPHYLP	1666				
Db	4187	VSDYMSAKEVND	SVIKKLDTPLA--	SKDAKRAVEN	QOATAP	IPN---	POPQNSMAQ	4241				
Qy	1667	PYLIRGV	PDTAALENRQTI	INDYITSQ	MEHNTATAMAQRA-----	-----	DMLR	1711				
Db	4242	ETAL-----	PTTSM	SVNNS---	NDHDTEDB-----	TETROL	PPAKPPIPTVGRPPGRGSSAKR	4291				
Qy	1712	GLSPRES-----	SLALN--	YAAGPR	GIIDLSQVP-----	HLPLV	VPPTPGTPA-----	1752				
Db	4292	GRQPRGAKVGG	PLNSVTAAP	PCVDSLV	VQPDNGVQTR	LRKPVTAPVTR	GRGKPPRN	4351				
Qy	1753	-----	TAMDR	LAYLPTAPOP	PTSSRHSS	PLPGGTHLTKPT	TTSSSRERDRDRER	1805				
Db	4352	LLLQOQLOQOQLD	IQRKGM	EMVTSATS	STPL---	PTPI	TSVLTAAEK--KARNOALT	4406				
Qy	1806	RDREKSI	LTSITTV	VEHAPIWR	PGTQSGSS	SGSSGGGSSSRP-----	-----	ASHSHAH	1858			
Db	4407	QAQEQNVASQV	GTGQDIYEF	HEDDGGE	PKPTISS	VAPSAEQRP	RLILITINKTOPSIK	4466				
Qy	1859	QHPIS	PRTDALQOR	PSVLHNTG-----	-----	MKIIT	AVEPSKPTVLR	1899				
Db	4467	NISEMEQTI	OOQOQOQSE	VISNTDPT	IGDNS	ESCNTKSRRL	QEKEDRSTVDIIEDVVR	4526				
Qy	1900	STSTSPVR	PAATPPP	PATHCPL	GTGLDGVYPTL	MEPVLL	PKEAPR-----	1944				
Db	4527	NTNT-----	PTGTGP--	HLPKAQTE-----	-----	PRSR	NAQAOKTDAVOIINA	4566				
Qy	1945	VARPERP-----	-----	ADTGH	FLAKPPARSGLL-----	-----	1970					
Db	4567	VGRPRSKOR	KRTIGEOTAN	LIEV	TASNATVAASHLA--	PPEGAG	VESHV	PQLDAKEVEPV	4625			
Qy	1971	----	PASSPKSGE	PRPLVP	PSVSHAT	IA	TPAKNLAPHAS	PDPPAPPASADPHREKT	2026			
Db	4626	SVUTPI	STPAPVSAAP	VTVPV-----	AMVPVK	PTMPQH--	PKKXIAAAE	IESQAIN	4678			
Qy	2027	QSKPF-----	SIQEL	ELRSLG	YHGSSY	PEGVEP	SVSPSL	THDK-----	GL	2070		
Db	4679	SSIPSGGL	PMHQTAAP	ATQKIT	TGGVAD	AVSKALVDP	VTGVTI	AGMPQCKEG	NLPAAATAA	4738		
Qy	2071	PKHLEELD-----	KSHLE	GLR	PKQP--	GPVK	LGEAAHL	PHLRPL	PESQSSSPLL	2120		
Db	4739	PANSSNED	GQAPPPQLOHQO	QOQHPQ	PPQOQAN	LQINTTL	IPSG	LNP	PI	ITALGKSVQL	4798	
Qy	2121	QTAPGV	KGHQ	RVVTLAQ--	HI	SEVIT	ODYTR-----	-----	HHQQL	SAPL	2159	
Db	4799	ETSAALLN	KPVS	VKGNAS	QVIT	QOQOQPI	VAPAKQPI	I	LQON	NPL	PTVLHHAQHTTVRP	4858
Qy	2160	PAPLYSP	PGASC	PVLDLRR-----	PPSD	LYLPP	PDHGP	ARGSPHS--	-----	EGGK	2204	

Db	4859	POPL	-----KAHVLNREKNIQQOLTFPTKQAVAPPPQH-----APHSGHMLLTDTAGNQ	4906
Qy	2205	RSPEPNKTSVLGGEGDIEPVSPPEGWTEPGHSR-----SAVYPLLVRDGEQTE	2253	
Db	4907	OLVQPOIIARHLOQQOHLQVNVPPPTAHSPHSPIRPSOQQOGLPGASISP-----QQQQ	4960	
Qy	2254	PSRMGSKSPGNTSQPPAPFASKLUTESAMVYKSKQBEINKLNTNRNEPEYNISQFTEI	2313	
Db	4961	PQTWVIKQAAASAQPQIL--HVVSXKSAVVPQPOQQ--QLPPTSSTGPHLQAKPNYSY	5015	
Qy	2314	F-----NMPAITGTGLM--TYRSQAVQEHASTNWNGLEAIRKA-----LMGK	2353	
Db	5016	APTVLPTFTLPAVQQOQQOHLQKONNQKGAQIQMPPHGIIMPTHGCMLLQOKLPAHLQPP	5075	
Qy	2354	YDQWESPPP-----LSANAFNPLNASASLPAAMPITTAADGRSDHFTLTSPGGGKA	2403	
Db	5076	QHQLNFSPPGKKNPVLHGLQSQIIMPVGVSGPPVSAVLKTAQQOVNSVVPVAGIRTA	5135	
Qy	2404	KVSGRESSRKAKSP-----APGLASGRPPSVSSVHSEGD--CNRRT-----PLTN	2447	
Db	5136	IPNISFQSQPRVSPVLVLPQGISG--VPPFDASLHDLGAVVSGRRTPSPPAHQOQASPITP	5193	
Qy	2448	RWVEDRPSAGSTFPFYNPLIMRLQAG-----VMASPPPPQL--PAGSGP---LAGPHAWD	2499	
Db	5194	NDSTYRGVTSRDFMLYQHHLMR--GGDYDDKMGSSPPLRLRRPGSGPPRTIAVPH----	5247	
Qy	2500	EEPKPLLCQYETLSDS	2516	
Db	5248	-----SLQSPQDRTAADS	5260	

RESULT 13

PCLO RAT

ID

PCLO RAT

STANDARD;

PRT;

5085 AA.

AC

Q9UKS6; Q9JUT1;

DT

28-FEB-2003 (Rel. 41, Created)

DT

28-FEB-2003 (Rel. 41, Last sequence update)

DT

10-OCT-2003 (Rel. 42, Last annotation update)

DE

Piccolo protein (Multidomain presynaptic cytomatrix protein).

GN

Rattus norvegicus (Rat).

OC

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX

NCBI_TaxID=10116;

[1]

SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH RABAC1.

RX

MEDLINE=20170257; PubMed=10707984;

RA

Fenster S.D., Chung W.J., Zhai R., Cases-Langhoff C., Voss B.,

RA

Garner A.M., Kaempf U., Kindler S., Gundelfinger E.D., Garner C.C.;

RT

"Piccolo, a presynaptic zinc finger protein structurally related to

RT

basoon";

RT

Neuron 25:203-214(2000).

[2]

SEQUENCE FROM N.A. (ISOFORM 1).

RA

Fenster S.D., Cases-Langhoff C., Gundelfinger E.D., Garner C.C.;

RL

Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

[3]

CALCIUM-BINDING ACTIVITY, AND MUTAGENESIS OF ASP-4668; ASP-4674;

RP

VAL-4688; MET-4689; VAL-4690; SER-4691; GLN-4692; ASN-4693 AND

RP

ALA-4694.

RX

MEDLINE=21181819; PubMed=11285225;

RA

Gerber S.H., Garcia J., Rizo J., Suedhof T.C.;

RT

"An unusual C(2)-domain in the active-zone protein piccolo;

RT

implications for Ca(2+) regulation of neurotransmitter release.";

RL

EMBO J. 20:1605-1619(2001).

CC

-I- FUNCTION: May act as a scaffolding protein involved in the

CC

organization of synaptic active zones and in synaptic vesicle

CC

trafficking (By similarity).

CC

-I- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin (By

CC

similarity).

CC

-I- SUBCELLULAR LOCATION: Concentrated at presynaptic side of synaptic

CC

junctions.

-1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9JKS6-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9JKS6-2; Sequence=VSP_003930, VSP_003931;
CC -1- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -1- SIMILARITY: Contains 2 C2 domains.
CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL; AF138789; AAF07822.2; -;
DR EMBL; AF227534; AAF63196.1; -;
DR HSSP; P04410; 1A25.
DR GO; GO:0045202; C:synaptic junction; IDA.
DR GO; GO:0005509; F:calcium ion binding; IDA.
DR GO; GO:0005544; F:calcium-dependent phospholipid binding; IDA.
DR GO; GO:0005522; F:profilin binding; ISS.
DR GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.
DR GO; GO:0016080; P:synaptic vesicle targeting; NAS.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR008899; Znf_piccolo.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF05715; ZF_piccolo; 2.
DR SMART; SM00239; C2; 2.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; 1.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
DR PROSITE; PS50106; PDZ; 1.
KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
Repeat; Alternative splicing.
FT DOMAIN 372 491
FT FT
FT ZN_FING 523 547
FT ZN_FING 1010 1033
FT DOMAIN 2351 2362
FT DOMAIN 4442 4536
FT DOMAIN 4653 4752
FT DOMAIN 4968 5059
FT VARSPPLIC 4876 4880
FT FT
FT VARSPPLIC 4881 5085
FT FT
FT MUTAGEN 4668 4668
FT FT
FT MUTAGEN 4674 4674
FT FT
FT MUTAGEN 4688 4688
FT FT
FT MUTAGEN 4688 4689
FT FT
FT MUTAGEN 4689 4689
FT MUTAGEN 4690 4691
FT FT
FT MUTAGEN 4692 4693
FT FT
FT MUTAGEN 4694 4694
FT FT
FT SEQUENCE 5085 AA; 552702 MW; 5A1BB543201A7450 CRC64;
SQ

```
Db 828 TPGTRPTAGQAAPPSSQPPQEQSRRFSLNGLGTTDAPKSSQ-----PTTPOETVTGKLF 883
Qy 1593 -----AKSPHSTVPEHHPH-----PISVEHLRLGVSGVDLYRSHIPLAFDPTSIPIR 1639
Db 884 GFGASIFSQANLSTAGQQPHQPTGPAAPSKQAPT-----PSQSPAAQGFPAKSTG 935
Qy 1640 GIP-----LDAAAAYLPRHLAPNPTYPHL-----YPPYLINGYDPTAALENR- 1682
Db 936 QLPPAPAKATAVKEAKAAENLESKEQAPTAKKTEKDKKPPPAKVKGKPPPEPEKAV 995
Qy 1683 -----QTIINDYITSOQ-MHNTATAMAQADMLRLGLSPR-----E 1717
Db 996 PAHRPDKTKPKPACPLCRTELN--LGSQEPNENFTCTCKNQVNCNLCGFNPTPHLTIQ 1053
Qy 1718 SSLALN-----YAAGPRGIIDLQVPHLPVLPPTGTPATAMDLAVLPTAPQPFSSRH 1772
Db 1054 EWLCLNCOTQRAISQILG-DWKG-----MPAPSPG-----KASPM----- 1089
Qy 1773 SSSPLSGPGLTHLTKPTTSSSERDRDRDREREKSIILTSITTTVEHAPIWRPOTE 1832
Db 1090 -----APAPSSQKPTGT-----QVKGKKEAGTEAEKPV----- 1122
Qy 1833 QSSSGSSGGGSSSRPASHAHQHSPISPRTQDALQORPSVLHNTGKGIITAVEP 1892
Db 1123 -----PEKETASIEKTPPMVTTDQKLE-----SEGKSKVSALPE 1158
Qy 1893 SKPT-VLRSTSTSPVR--PAATPPATHCPGLGTLGTVYPTLMEPVLLPKEAPRVARPE 1949
Db 1159 KKPSEEEKAIADKERRKPPAEKPLEE-----KKPIPVDKKLPEAKPL 1204
Qy 1950 RPRADTGHAFI-----AKPPARSGLEFPASSPKSGSPR 1982
Db 1205 SSEGEKEHILKAHVQIPEEBPTGKVAAGKEEQPDSRPEALPGATFLTLPRAGEKER 1264
Qy 1983 PLVPPVS-----GHATARTPAKULAPHASDPDPAPASAD----- 2020
Db 1265 AVAQPQABSGSKDGSGERSKEKTEEDKSDTSSSQPKSPQLSDTGYSDGSGSLGE 1324
Qy 2021 -----PHREKTSQFPFSIQEELRSLGVHSGSYSPGVEPVSPVSPSLTHDKGLPKHLE 2075
Db 1325 IPSLIPSEKDLK-----GLKDSFQES-----SPSSPSDLAKLESTVLSIL 1368
Qy 2076 ELDKSHLEGLRKPQPGVKLGGEAAHLPHLRPLPESQSPSSPLIQTAPGVKGHQRVVTL 2135
Db 1369 EAQASTLVGERAKKTPQKISPEK-----PDQ-----QKTQTA 1403
Qy 2136 AQHISEVITQYTRHHPOOLSAPLPAPLYSPGASCPVLDLRRPPSDLYLPPDHGAPAR 2195
Db 1404 SETDLITISEBEIKESQEKVSPKXDSQGFPSRK-----EHKEKPELV-----DQLSPR 1454
Qy 2196 GSPHSEGGKSPENKTSVLGGEDGIEPVSPPEGMTEPGHSRAVYVPLLYRDGEQTEPS 2255
Db 1455 AS-----YDSVESSESENSEPVVRKRRTSIGSSSDSEYKQEDSQSGSEED 1501
Qy 2256 -----RMGSKSPGNTSQPPAFSKLTESNMAVSKKQB-----INNK 2293
Db 1502 FIRKQIEMSADEASGSEDEFIRSQLKETISGVGESQKREAEKKGKGVAGKHRLTRK 1561
Qy 2294 LNT-----HNRNE-----PEYNI-----SQGTIFNMPALTGTGLMYRSQA 2331
Db 1562 SSTSFDDDAGRHRSHWDEDDTFDBSPKFKRETQSQESELV---VAGGGLRFRFKTIE 1618
Qy 2332 VQEHASTNMGLEAIRKALMGKYDOWERSPLSANAF-----NPLNASASLP 2378
Db 1619 LNSTIADKYSESSQKKTIL-----YFDEBELEMESLTDSPEDRSRGSGSSLSHASTP 1674
Qy 2379 AAMP--ITAAGRSDDHTLTSPGGGKAKVSGRPSRKAAPSAPG 2420
Db 1675 GTSPTSVSLSDESDSFSHKKGESK-----QQRKARHSHG 1711
```

RESULT 14

```
MAPA HUMAN
ID AC MAPA HUMAN STANDARD; PRT; 2805 AA.
AD P78559; Q95643; Q12973; Q15882; Q9UUT4;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1A (MAP 1A) (Proliferation-related
protein p80) [Contains: MAP1 light chain LC2].
GN MAP1A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OK NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97001161; PubMed=8812494;
RA Fink J.K., Jones S.M., Esposito C., Wilkowiak J.;
RT "Human microtubule-associated protein 1a (MAP1A) gene: genomic
organization, cDNA sequence, and developmental- and tissue-specific
expression."
RL Genomics 35:577-585(1996).
RN [2]
RP SEQUENCE OF 1-1825 FROM N.A.
RC TISSUE=Ovarian carcinoma;
RA Chen Z.C., Fadiel A., Naftolin F.;
RT "Identification of a novel protein (p80) in ovarian carcinoma cells."
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 78-1687 FROM N.A.
RC TISSUE=Brain;
RA Ohtani K., Rutherford T., Sakamoto H., Naftolin F.;
RT "Microtubule associated protein 1A (MAP1A) in human brain - DNA
sequence and physiological role."
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 134-419 FROM N.A.
RC TISSUE=Fetal muscle;
RA Chiamnikulchai N., Pasturaud P., Richard I., Auffray C.,
RL Beckmann J.S.;
RN [5]
RP SEQUENCE OF 1607-1883 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95356255; PubMed=7629894;
RA Fukuyama R., Rapoport S.I.;
RT "Brain-specific expression of human microtubule-associated protein 1A
(MAP1A) gene and its assignment to human chromosome 15."
RL J. Neurosci. Res. 40:820-825(1995).
CC -!- FUNCTION: Structural protein involved in the filamentous cross-
bridging between microtubules and other skeletal elements.
CC -!- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
with MAP1A and MAP1B proteins.
CC -!- TISSUE SPECIFICITY: Brain.
CC -!- DOMAIN: The basic region containing the repeats may be responsible
for the binding of MAP1A to microtubules.
CC -!- PTM: Various serine residues may be phosphorylated by cAMP kinase.
CC -!- PTM: LC2 IS COEXRESSED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
BOTH MAP1A AND MAP1B.
CC -!- SIMILARITY: TO MAP1B.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/
or send an email to license@isb-sib.ch).
-----
EMBL; U38291; AAB41132.1; -
EMBL; U38292; AAB41133.1; -
EMBL; AF200415; AAF08305.2; -
EMBL; U80458; AAD00355.1; -
```


3292	PGATPTMSQTFPMVQQLHQHQHTTVISGHTSPVRMPSLPGWQPNAPAHPLNPRI	3351	RA	Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,		
2013	-----APPASADPHREKTSQKPF-----	2032	RA	Wohldman P.;		
3352	OPPIAQLPIKCTPAGTVSNANPQSGPPRVEFDNNPFSEGFQERKERLREQERQ	3411	RT	"2.2 Mb of contiguous nucleotide sequence from chromosome III of C.		
2033	-----IOELRLSLGHGSSVSP-----	2061	RL	<i>elegans</i> .";		
3412	RIQLMQEVDQRALQQRMEHQMGVGEISSSRSTVSQIPFYSSDLPFCDFMPLGLPQ	3471	RN	Nature 368:32-38(1994).		
2062	PSLTHDKLGLPKHLSEL-----DKSHLEGELRP-----	2104	RP	[2]		
3472	S-----POHQOQGVQLQQNIQQSINSPTQTFMQTNERQVGPSPFVDPSPSIP	3523	RA	REVISIONS, AND ALTERNATIVE SPLICING.		
2105	HLRLPES-----QPSSSPLLTQAPGVKGHQRVVTLLAQHISEVITQD	2146	RP	Waterston R.;		
3524	VGSFNFSSVKQHGKNSLCTSFQSQSVRPSFTPALPAAPPV-----ANSLP	3573	RL	Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.		
2147	YTRHHPQQLSAPLPAPLYSPGASCPVLDLRPPSDLYLPPDHGAPARGSPHSEGGKRS	2206	CC	- - SUBCELLULAR LOCATION: Nuclear (Probable).		
3574	STITHG-----HSYFGSTQSLIQLY-----SDII-----PEKGGKKR	3606	CC	- - ALTERNATIVE PRODUCTS:		
2207	PEPNTSVLGGEDGIEPVSPPEGMT-----EPGHSRAVYLLYRDGEQTEPSRMGSKSPG	2263	CC	Event=Alternative splicing; Named isoforms=2;		
3607	TRKKQRD-----DDAESTKAPSTHSDITAPPTFCISETTSTPAVSPSELPPQADQESVEPV	3664	CC	Name=b;		
2264	NTSQPPAFPSKL-TESNSAMVSKQKQINKLNTNHRNEPEYNISQPTETFNMPAITGT	2322	CC	isoId=p34333-1; Sequence=Displayed;		
3665	GFSTPNMAAGQLCTELEKLP-----NSDFSQATFNQQTAYANSEVDKLSMETPAKT--	3715	CC	isoId=p34333-2; Sequence=VSP 003306, VSP 003307;		
2323	GLMTRSAQVQEHASTNMGLEAIRKALMGKYDQWEESSPPLSANAFNPLNASASLPAAMP	2382	CC	Note=No experimental confirmation available;		
3716	-----EEIKLEKAETESCPG-----QEPKLEEQNGSKVEGNA-----VACP	3752	CC	- - SIMILARITY: Contains 1 Myb-like domain.		
2383	ITAADGRSDHTLTSP-----GGGKAKVSGRPSRRKAKSPAGLASGRPPSPVSVHSEGDC	2439	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
3753	VSAQS-PHSAQAPAAKGDGNSNELLKHLKLNKSS-----LLNQKPE--GSIQSDDC	3804	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -		
2440	NRTPTNVRWEDRPSASGTFPPYNNPLMRLOAGVMASPPPPGLPAGS-----	2488	CC	the European Bioinformatics Institute. There are no restrictions on its		
3805	TKDNKL---VEKQNPAGLQT-----LGAQMGQFGCGNQLPKTDGSETKKQSKRTQ	3855	CC	use by non-profit institutions as long as its content is in no way		
2489	-GPLAGPH---HAWDEEPKPLLCQYETLS 2514		CC	modified and this statement is not removed. Usage by and for commercial		
3856	RTGEKAAPSKKKRKKDEBEKQAMYSSTDFT 3886		CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).		
RESULT 17					EMBL; L15188; AAA27949.3; -	
YK26 CAEL					EMBL; L15188; SAA27950.1; -	
ID	YK26	CAEL	STANDARD;	PRT;	1780	AA.
AC	P34333;	P34332;				
DT	01-FEB-1994	(Rel. 28, Created)				
DT	28-FEB-2003	(Rel. 41, Last sequence update)				
DT	28-FEB-2003	(Rel. 41, Last annotation update)				
DE	Hypothetical protein C14B9.6 in chromosome III.					
GN	C14B9.6/c14B9.5					
OS	Caenorhabditis elegans.					
OC	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;					
OC	Rhabditidae; Peloderinae; Caenorhabditis.					
OX	NCBI_TaxID=6239;					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN=Bristol N2;					
RX	MEDLINE=94150718; PubMed=7906398;					
RA	Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,					
RA	Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,					
RA	Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,					
RA	Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,					
RA	Johnston L., Jones M., Karshaw J., Kirsten J., Laister N.,					
RA	Lafrelle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,					
RA	Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,					
RA	Sims M., Smaldon N., Smith A., Smith M., Sonhammer E., Staden R.,					
RA	Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,					

Qy	270	IKINQAMRKLLI-----LVFKRNHARKQWKQFCQRYDQLEALEKKVERIENN	319	QY	
Db	66	LKVNGLDLLIFCSFCLFVLVFTFKELARE-----YDEKFAQDDI-QKWEET	116	Db	
Qy	320	PRRAKESKREYVEKQFPIRKQELQRMQSRVGQSGLSMSAARSEHEVSEIDGL	379	QY	
Db	117	SERKEYAEFHRVQAESEFPFKRERERW-----AAAEIRGEDEKSM	161	Db	
Qy	380	SEQUENLEKQMR-QLAVIPPMLYDADQORIKFINNGLMADPMKYKQOVNM--WSEQ	435	QY	
Db	162	LAKEHADKKIRLGVAKIPRLITSESKQWDFEVRPGSLKDMKKHRSQVLDRLSEWSP	221	Db	
Qy	436	EKETREKFMQHPKNGFIAGFLERKTVAEVCVLYITTKNKNYKSLVRSYRRGKSQ	495	QY	
Db	222	ERSLFKSQADHVKIFHGLTFFVDKTDASDLVLFYNNKTKEDYKDKPKKRVTKYKV	281	Db	
Qy	496	QQQQQQQQQQQQQQQQMPRSSQEEK-----DEKEKEAEK-----EE	533	QY	
Db	282	APPSVEELAYFRMPPPLDFSSFPKNSLMCYFCCTRWGIDLNGTTPMPKEAYEIFA	341	Db	

RESULT 17

YKZ6 CAEEL

ID_YKZ6 CAEEL STANDARD; PRT; 1780 AA.

AC P34333; P34332; AC

DT 01-FEB-1994 (Rel. 28, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein C14B9.6 in chromosome III.

GN C14B9.6/C14B9.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditoidea; Rhabditida; Rhabditidae; Pelodérinae; Caenorhabditis.

OC Rhabditidae; Pelodérinae; Caenorhabditis.

OX NCBI_TaxID=6239;

UN [1]

RN SEQUENCE FROM N.A.

RP STRAIN=Bristol N2;

RC MEDLINE=94150718; PubMed=7906398;

RX Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,

RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,

RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,

RA Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,

RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,

RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,

RA Sims M., Smaildon N., Smith A., Smith M., Sonnhammer E., Staden R.,

RA Suleston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,

RP TISSUE SPECIFICITY, AND INTERACTION WITH PROFILIN.
RC TISSUE=Brain;
RX MEDLINE=9439764; PubMed=1050862;
RA Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,
RA Kilimann M.W.;
RT "Aczonin, a 550-kd putative scaffolding protein of presynaptic active
RT zones, shares homology regions with rim and bassoon and binds
RT profilin";
RL J. Cell Biol. 147:151-162(1999).
RN [2]
RP REVISIONS.
RA Kilimann M.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 4502-4682 FROM N.A.
RC TISSUE=Brain;
RA Huang W., Jin W., Huang C., Chen B., Zhang J., Ju G.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP INTERACTION WITH RIMS2.
RX MEDLINE=22384373; PubMed=12401793;
RA Fujimoto K., Shibasaki T., Yokoi N., Kashima Y., Matsumoto M.,
RA Sasaki T., Tajima N., Iwanaga T., Seino S.;
RT "Piccolo, a Ca2+ sensor in pancreatic beta-cells. Involvement of
RT CAMP-GEFII.Rim2.Piccolo complex in cAMP-dependent exocytosis.";
RT J. Biol. Chem. 277:50497-50502(2002).
CC -!- FUNCTION: May act as a scaffolding protein involved in the
CC organization of synaptic active zones and in synaptic vesicle
CC trafficking.
CC -!- SUBUNIT: Interacts with Rabac1/Pral, RIMS2 and profilin.
CC -!- SUBCELLULAR LOCATION: Concentrated at the presynaptic side of
CC synaptic junctions.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q9QYX7-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9QYX7-2; Sequence=VSP_003929; VSP_003929;
CC -!- TISSUE SPECIFICITY: Highly expressed in brain. Low levels found in
CC stomach. Not detected in other tissues analyzed including adrenal
CC gland, testis and pancreas.
CC -!- DOMAIN: C2 domain 1 is involved in binding calcium and
CC phospholipids. Calcium binds with low affinity but with high
CC specificity and induces a large conformational change.
CC -!- SIMILARITY: Contains 2 C2 domains.
CC -!- SIMILARITY: Contains 1 PDZ/DHR domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC ENBL; Y19185; CAB60731.2; --
CC ENBL; Y19186; CAB60732.2; --
CC ENBL; AF181269; AAD55786.2; --
CC HSP; P04410; 1A25.
CC MGD; MGI:1349390; Pclo.
CC GO; GO:0045202; C:synaptic junction; IDA.
CC GO; GO:0005509; F:calcium ion binding; ISS.
CC GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.
CC GO; GO:0005522; F:profilin binding; IDA.
CC GO; GO:0019933; P:CAMP-mediated signaling; IDA.
CC GO; GO:0007010; P:cytoskeleton organization and biogenesis; IDA.
CC GO; GO:0030073; P:insulin secretion; IDA.
CC GO; GO:0017157; P:regulation of exocytosis; IDA.
CC GO; GO:0016080; P:synaptic vesicle targeting; NAS.
CC InterPro; IPR000008; C2.
CC InterPro; IPR001478; PDZ.
CC InterPro; IPR008899; Znf_piccolo.
CC Pfam; PF00168; C2; 2.

DR Pfam; PF00595; PDZ; 1.
DR Pfam; PF03715; Zf_piccolo; 2.
DR SMART; SM00239; C2; 1.
DR SMART; SM00228; PDZ; 1.
DR PROSITE; PS00499; C2 DOMAIN 1; 1.
DR PROSITE; PS00004; C2 DOMAIN 2; 2.
DR PROSITE; PS01006; PDZ; 1.
KW Calcium/phospholipid-binding; Metal-binding; Zinc; Zinc-finger;
KW Repeat; Alternative splicing.
FT DOMAIN 371 470
FT 10 X 10 AA TANDEM APPROXIMATE REPEATS OF
FT P-A-K-P-Q-P-Q-P-X.
FT ZN_FING 502 526
FT C4-TYPE (POTENTIAL).
FT ZN_FING 967 990
FT C4-TYPE (POTENTIAL).
FT DOMAIN 2305 2329
FT POLY-PRO.
FT DOMAIN 4394 4488
FT PDZ.
FT DOMAIN 4607 4705
FT C2 DOMAIN 1.
FT DOMAIN 4922 5012
FT C2 DOMAIN 2.
FT TRPTN -> SKRRK (in isoform 2).
FT /FTID=VSP_003928.
FT Missing (in isoform 2).
FT /FTID=VSP_003929.
FT VARSPLIC 4834 5038
FT VARSPLIC 5038 AA; 547600 MM; DADA460CF3B40888 CRC64;
FT SEQUENCE 5038 AA; 547600 MM; DADA460CF3B40888 CRC64;
Query Match 3.3%; Score 434.5; DB 1; Length 5038;
Best Local Similarity 19.8%; Pred. No. 8.6e-07; Indels 663; Gaps 91;
Matches 391; Conservative 216; Mismatches 701;
688 AAASEEAPPPVVEDEMEASGVSGNEEMVEAEALHASGNEVPRGECSPATVNNSSD 747
23 AGSGGALHPGIPAGMEADLSQLS---EEERQIAVMSRAQGLPKG--SVPAAMAE-- 75
748 TESIPSPHTEAAKDTGQNGPKPATILGADGPPGPPPTPRR-TSRAPITPTASEATGAP 806
76 ----PSMHRKQELDSQ-APQQP-----GKPPDPGRFPQHGLSKSRITDTFRSE---OK 121
807 TPPAPPSPAPPPVVPVKEEKEETAAAPVEBG--EEOKPPAAEELAV-----DTGKA 858
122 LPCRSPSTISLKSKSRITDPKEEYKSMMP---GFFSEVNPLSAVSVVKNFNPDLISD 178
859 EEPVKSCTEAELEGPAKGDAEAAEATAFGALKAE-----KKEGSGRATTAKSS 909
179 SEA/QEETK-----KQVAKDQKSGEGITKPSLQQPSPKLIPKQQPGKEV----- 226
910 GAPDSSSATCSADEVDEAGGDKNLLSP---RSLLTPTGDPANASP--QKPLDLK 964
227 -IPDIPSKSVSS-----QAEKTKPQAPGTAKFSQQSPAQTPAQAQKFAVQAQGPAP 278
965 -QLKQRAAALPIPIQVTKVHEPPREDAAPTKAPAPPQPQNLOPESDAPQPGSS----- 1018
279 ATVQQPGPAKSPAQ-----PAGTKGSPAQPPVTAKPQAQAGLEKTSLQQPGPKSLAQ 332
1019 -----PRGKSRSPAPPADKEAFAE-----AQKLPG--DPPCWTSLGLPPVP----- 1058
333 PGQGVPPGPAKSPAQPGTAKLPAQPGPQTASKVFGTKTTPAQLSG-PKPTPAQPGP 391
1059 ----PREVIKASHPAPPSAFSYP-----PGHPLPLGLHDTARVPLPRE--PTIS 1103
392 TKPSFPQPIPAKPOQPPVATKPOQPPAPAKPOQHPTPA----KQPQQPTPAKPOQ 447
1104 NPPPLISSAKHPSV-----LERQIGAIISQMSVQ-LHVPYSEHAKAPGVPTMGLPLMDP 1158
448 QPTPAKPOQHPLGKPSAQPSKISQTVTGRPLQAPPTSAQAQ-----AQGL-----S 498
1159 KKLAPFSGVKQEQLSPRQAGPPPSLGVPTAQEASVLGALSGVPGSGITK----- 1210
499 KTCPLCNTTELLH-----TPEKANFNTCTEQSTVCSLCGNFNPPLHTEIKEWLCLN 552
1211 -----GIPSTRVPSDSAITYRGSITHTGTPADVLYKGTITRIIGSDSPSRLDRG 1258
553 CQMQRALGGELAAIPSSQPTPKAASVQPATASKSPV-----PSQAASP 596
1259 REDSLPKGHVIYEGKKHVLSEYEGMSTVQCSKEDGRSSSGPPHETAAPKRTYDMWEGRV 1318
::: |||

Db 597 KKE-LP-----SKQD-----SPK-----608
Qy 1319 GRAISSASIEGLMGRAIPERHSHLKEQHHIRGSITQGIIPRSYVEAQEDYLREAKLL 1378
Db 609 -----APESKPPPLVKQPTLHGPTATAPQPPV-----637
Qy 1379 KREGTPPPPSRDLTEAYKTOALGPL-KLPAHEGLVATVKEAGRSIHEIPREELRHTP 1437
Db 638 -AALPKPAPPKP-SAALPEQAKAFVADVEPKQPTTETLTDSPSSAAATSKPAI-----691
Qy 1438 ELPLAPRLKGSITQGTPLKYDTGASTTGSKKHVDVRSGLIGPGTTPPVHPLDVMADAR 1497
Db 692 ---LSSQVQAQAQVTTAPLTKDTSKTS-----QSFPTGDTTTPDLSK 732
Qy 1498 ALERACYBESLSRGTSSSGSSTARCAPVIVPELGPQRSPLTYEDHGAPFAGHLPRG 1557
Db 733 AMPRASDSKIVSHGPTSESKD-----PVQKKEPKKAQTKVTPKPTDKP-----VPGK 782
Qy 1558 SPV-----TWREPTPRLOEGLSSKASQDKLTSTPREIAKSPHSTVPEHHPIPTSPY 1611
Db 783 SPTSPGTRPTTGOATPOSQO---PPKPPQSRFRSLNLGGIADAPKS-----QPTTPQ 832
Qy 1612 EHLRGVSGVDLYRSHIPLADPTSI PRGIPLDAAAAYLPHLAPNTYPHLYPPYLIR 1671
Db 833 ETVTGKLP-----FGAIFSOASNLISAGQOAP-HPQTGPAAPSKOAP-----876
Qy 1672 GYPDTAALENQTIINDYITSQOMHHNTATAMORADMLRGLSPRESSIALNYAAGPRGI 1731
Db 877 -----PPSQTLA---AQGP---887
Qy 1732 IDLSQVPHLVLPPTGPTATAMDRLAYLTPAQPFSSRHSPLSPGPGTHLTKPTTT 1791
Db 888 -----PKSTG-----QHPSAKATKAVKTKG 910
Qy 1792 SSSERDRDRDREREKSLTSTTVEHAPIWRGTQSSGSSSSSGGGSSSRP 1851
Db 911 PAENLEAKPAQPTVKAEKD-----KKHPP-----GKVS KP 943
Qy 1852 ASHSHAHQHSIPRTQDALQORSVLHNTGMKGIIITAVEPSKPTVLKSTSTSPVPPAA 1911
Db 944 -----PTEPE-KAVLAQKPDK-----TTXPKPACPLCRTELNVGQDPP-- 982
Qy 1912 TFPATHCP-IGGTLGDVYPT-----LMPEVLLPKAPRVARPERPRADTGHAFKAPPA 1965
Db 983 NFNTCTECKNQVCLNGFNPTPHLTEIQEWCLNCQTORAISGQLGMD-----KNPPA 1036
Qy 1966 RSGLEPASSPKG-SERPPLVPVSGHA-----TIARTPAKNLAPHASPD-PPAPPAS 2017
Db 1037 SSG--PKASPVPAFAEPFPQKPTTAAHAHAKKKKXTEVKAETEKIPEKETESIEKTPPAV 1094
Qy 2018 ASDPHREKTQSKPFSIQELELSLGYHGSSYSPEGVPEVPVSPVPSLTHDKGLPKHLEL 2077
Db 1095 ATD-----QKLE-----ESEVTKSLVSLVLEKXPS-EEKALPADKKE- 1131
Qy 2078 DKSHLELRLPKQPGVKVKGGAHLPHRLPLPESQ-----PSSPLLQATPCVKGHQ 2130
Db 1132 -----KXPP-----AAEAPPLEBKPIPDQKLPPDAKPSASGEKRDLLKAHV 1176
Qy 2131 RVV-----TLAQHISEVITQDTRHHPOQLSAPLPAPLYFPGASCPLV--DLRRPPSD 2182
Db 1177 QIPEEGPIGVKSLACEGEOQPDTR--PEDL-----PGATPQTLPKDRQKESRD 1223
Qy 2183 LYLPPPDHGA-PARGSPHSEGGKSPENKTSVLGGGEDGIEPVSPPEGMTEPGHRSASV 2241
Db 1224 VTQPAEGTAGEGGEFSKORTEKEEDKSDTS-----SSQPKS-FOGLSDTGYSSDGI 1276
Qy 2242 -----YPLLVRDGE-----2250
Db 1277 SGSLGEIPLSPIDEKLLGLKKDSSFQESSPSSPSDLAKLESTVLISLEAQASTLVGE 1336
Qy 2251 -----QTPSRMGSKSPGN--TSQPAFPKSLTENSAMVSKKEINKLNT-----2296
Db 1337 KAEEKTQPKVSPQPODQKQTOTPTSETRDISIEEIKESQEKVKVTSKKDSQAQGFPSRK 1396

Qy 2297 HNRNEPE-----YNISQFGTEIFNMPAI---TGTGLMYRSQAVQEHASTNMG 2342
Db 1397 EHKENPELVDDLSPPRASDYSDVEDSESESPVARRKRTSTIGSSSSSEYKQEDSQSGE 1456
Qy 2343 -EAIIRKALMKYQOWEESPPLSANAFNPLNASASLPAAMPITADGRSDHLLTSPGGGG 2401
Db 1457 DEDFIRKQII-----EMSADE-----DASGS-----EDEFIRSQLKEIGV 1493
Qy 2402 KAKVSGRPSSRKAKSPAGLASGRPPSVSVHSEGDGNNRRTPLTNRWED 2452
Db 1494 TESQKRETKGKGPSA---GKHRLTRKSTSTFDDAGRR-----HSWHD 1536
RESULT 19
SHK1 RAT
ID _SHK1 RAT STANDARD; PRT; 2167 AA.
AC Q9WV48; Q9QZ28; Q9WU13; Q9WUE8;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP
interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
interacting protein) (SSTR interacting protein) (SSTRIP).
GN SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
DLG4.
RC TISSUE=Brain;
RX MEDLINE=99419021; PubMed=10488079;
RA Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "Synamon, a novel neuronal protein interacting with synapse-associated
protein 90/postsynaptic density-95-associated protein.";
RL J. Biol. Chem. 274:27463-27466(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=20549637; PubMed=10958799;
RA Naebitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,
RA Weinberg R.J., Worley P.F., Sheng M.;
RT "Shank, a novel family of postsynaptic density proteins that binds to
the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RL Neuron 23:569-582(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobaben S., Suedhof T.C., Stahl B.;
RT "The G protein-coupled receptor C1i interacts directly with proteins
of the shank family.";
RL J. Biol. Chem. 275:36204-36210(2000).
RN [4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
DEVELOPMENTAL STAGE.
RC TISSUE=Brain;
RX MEDLINE=99436166; PubMed=10506216;
RA Lim S., Naebitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
RT "Characterization of the Shank family of synaptic proteins. Multiple
genes, alternative splicing, and differential expression in brain and
development.";
RL J. Biol. Chem. 274:29510-29518(1999).
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
multidomain proteins present in human and rodent brain.";
RL J. Biol. Chem. 274:32997-33001(1999).

[6] INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
MEDLINE=99360651; PubMed=10433269;
Tu J.C., Xiao B., Naibitt S., Yuan J.P., Petralia R.S., Brakeman P.,
RA Deane A., Aakalu V.K., Lanthan A.A., Sheng M., Worley P.F.;
RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
RT postsynaptic density proteins.";
RL Neuron 23:583-592(1999).
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[22]
[23]
[24]
[25]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[34]
[35]
[36]
[37]
[38]
[39]
[40]
[41]
[42]
[43]
[44]
[45]
[46]
[47]
[48]
[49]
[50]
[51]
[52]
[53]
[54]
[55]
[56]
[57]
[58]
[59]
[60]
[61]
[62]
[63]
[64]
[65]
[66]
[67]
[68]
[69]
[70]
[71]
[72]
[73]
[74]
[75]
[76]
[77]
[78]
[79]
[80]
[81]
[82]
[83]
[84]
[85]
[86]
[87]
[88]
[89]
[90]
[91]
[92]
[93]
[94]
[95]
[96]
[97]
[98]
[99]
[100]
[101]
[102]
[103]
[104]
[105]
[106]
[107]
[108]
[109]
[110]
[111]
[112]
[113]
[114]
[115]
[116]
[117]
[118]
[119]
[120]
[121]
[122]
[123]
[124]
[125]
[126]
[127]
[128]
[129]
[130]
[131]
[132]
[133]
[134]
[135]
[136]
[137]
[138]
[139]
[140]
[141]
[142]
[143]
[144]
[145]
[146]
[147]
[148]
[149]
[150]
[151]
[152]
[153]
[154]
[155]
[156]
[157]
[158]
[159]
[160]
[161]
[162]
[163]
[164]
[165]
[166]
[167]
[168]
[169]
[170]
[171]
[172]
[173]
[174]
[175]
[176]
[177]
[178]
[179]
[180]
[181]
[182]
[183]
[184]
[185]
[186]
[187]
[188]
[189]
[190]
[191]
[192]
[193]
[194]
[195]
[196]
[197]
[198]
[199]
[200]
[201]
[202]
[203]
[204]
[205]
[206]
[207]
[208]
[209]
[210]
[211]
[212]
[213]
[214]
[215]
[216]
[217]
[218]
[219]
[220]
[221]
[222]
[223]
[224]
[225]
[226]
[227]
[228]
[229]
[230]
[231]
[232]
[233]
[234]
[235]
[236]
[237]
[238]
[239]
[240]
[241]
[242]
[243]
[244]
[245]
[246]
[247]
[248]
[249]
[250]
[251]
[252]
[253]
[254]
[255]
[256]
[257]
[258]
[259]
[260]
[261]
[262]
[263]
[264]
[265]
[266]
[267]
[268]
[269]
[270]
[271]
[272]
[273]
[274]
[275]
[276]
[277]
[278]
[279]
[280]
[281]
[282]
[283]
[284]
[285]
[286]
[287]
[288]
[289]
[290]
[291]
[292]
[293]
[294]
[295]
[296]
[297]
[298]
[299]
[300]
[301]
[302]
[303]
[304]
[305]
[306]
[307]
[308]
[309]
[310]
[311]
[312]
[313]
[314]
[315]
[316]
[317]
[318]
[319]
[320]
[321]
[322]
[323]
[324]
[325]
[326]
[327]
[328]
[329]
[330]
[331]
[332]
[333]
[334]
[335]
[336]
[337]
[338]
[339]
[340]
[341]
[342]
[343]
[344]
[345]
[346]
[347]
[348]
[349]
[350]
[351]
[352]
[353]
[354]
[355]
[356]
[357]
[358]
[359]
[360]
[361]
[362]
[363]
[364]
[365]
[366]
[367]
[368]
[369]
[370]
[371]
[372]
[373]
[374]
[375]
[376]
[377]
[378]
[379]
[380]
[381]
[382]
[383]
[384]
[385]
[386]
[387]
[388]
[389]
[390]
[391]
[392]
[393]
[394]
[395]
[396]
[397]
[398]
[399]
[400]
[401]
[402]
[403]
[404]
[405]
[406]
[407]
[408]
[409]
[410]
[411]
[412]
[413]
[414]
[415]
[416]
[417]
[418]
[419]
[420]
[421]
[422]
[423]
[424]
[425]
[426]
[427]
[428]
[429]
[430]
[431]
[432]
[433]
[434]
[435]
[436]
[437]
[438]
[439]
[440]
[441]
[442]
[443]
[444]
[445]
[446]
[447]
[448]
[449]
[450]
[451]
[452]
[453]
[454]
[455]
[456]
[457]
[458]
[459]
[460]
[461]
[462]
[463]
[464]
[465]
[466]
[467]
[468]
[469]
[470]
[471]
[472]
[473]
[474]
[475]
[476]
[477]
[478]
[479]
[480]
[481]
[482]
[483]
[484]
[485]
[486]
[487]
[488]
[489]
[490]
[491]
[492]
[493]
[494]
[495]
[496]
[497]
[498]
[499]
[500]
[501]
[502]
[503]
[504]
[505]
[506]
[507]
[508]
[509]
[510]
[511]
[512]
[513]
[514]
[515]
[516]
[517]
[518]
[519]
[520]
[521]
[522]
[523]
[524]
[525]
[526]
[527]
[528]
[529]
[530]
[531]
[532]
[533]
[534]
[535]
[536]
[537]
[538]
[539]
[540]
[541]
[542]
[543]
[544]
[545]
[546]
[547]
[548]
[549]
[550]
[551]
[552]
[553]
[554]
[555]
[556]
[557]
[558]
[559]
[560]
[561]
[562]
[563]
[564]
[565]
[566]
[567]
[568]
[569]
[570]
[571]
[572]
[573]
[574]
[575]
[576]
[577]
[578]
[579]
[580]
[581]
[582]
[583]
[584]
[585]
[586]
[587]
[588]
[589]
[590]
[591]
[592]
[593]
[594]
[595]
[596]
[597]
[598]
[599]
[600]
[601]
[602]
[603]
[604]
[605]
[606]
[607]
[608]
[609]
[610]
[611]
[612]
[613]
[614]
[615]
[616]
[617]
[618]
[619]
[620]
[621]
[622]
[623]
[624]
[625]
[626]
[627]
[628]
[629]
[630]
[631]
[632]
[633]
[634]
[635]
[636]
[637]
[638]
[639]
[640]
[641]
[642]
[643]
[644]
[645]
[646]
[647]
[648]
[649]
[650]
[651]
[652]
[653]
[654]
[655]
[656]
[657]
[658]
[659]
[660]
[661]
[662]
[663]
[664]
[665]
[666]
[667]
[668]
[669]
[670]
[671]
[672]
[673]
[674]
[675]
[676]
[677]
[678]
[679]
[680]
[681]
[682]
[683]
[684]
[685]
[686]
[687]
[688]
[689]
[690]
[691]
[692]
[693]
[694]
[695]
[696]
[697]
[698]
[699]
[700]
[701]
[702]
[703]
[704]
[705]
[706]
[707]
[708]
[709]
[710]
[711]
[712]
[713]
[714]
[715]
[716]
[717]
[718]
[719]
[720]
[721]
[722]
[723]
[724]
[725]
[726]
[727]
[728]
[729]
[730]
[731]
[732]
[733]
[734]
[735]
[736]
[737]
[738]
[739]
[740]
[741]
[742]
[743]
[744]
[745]
[746]
[747]
[748]
[749]
[750]
[751]
[752]
[753]
[754]
[755]
[756]
[757]
[758]
[759]
[760]
[761]
[762]
[763]
[764]
[765]
[766]
[767]
[768]
[769]
[770]
[771]
[772]
[773]
[774]
[775]
[776]
[777]
[778]
[779]
[780]
[781]
[782]
[783]
[784]
[785]
[786]
[787]
[788]
[789]
[790]
[791]
[792]
[793]
[794]
[795]
[796]
[797]
[798]
[799]
[800]
[801]
[802]
[803]
[804]
[805]
[806]
[807]
[808]
[809]
[810]
[811]
[812]
[813]
[814]
[815]
[816]
[817]
[818]
[819]
[820]
[821]
[822]
[823]
[824]
[825]
[826]
[827]
[828]
[829]
[830]
[831]
[832]
[833]
[834]
[835]
[836]
[837]
[838]
[839]
[840]
[841]
[842]
[843]
[844]
[845]
[846]
[847]
[848]
[849]
[850]
[851]
[852]
[853]
[854]
[855]
[856]
[857]
[858]
[859]
[860]
[861]
[862]
[863]
[864]
[865]
[866]
[867]
[868]
[869]
[870]
[871]
[872]
[873]
[874]
[875]
[876]
[877]
[878]
[879]
[880]
[881]
[882]
[883]
[884]
[885]
[886]
[887]
[888]
[889]
[890]
[891]
[892]
[893]
[894]
[895]
[896]
[897]
[898]
[899]
[900]
[901]
[902]
[903]
[904]
[905]
[906]
[907]
[908]
[909]
[910]
[911]
[912]
[913]
[914]
[915]
[916]
[917]
[918]
[919]
[920]
[921]
[922]
[923]
[924]
[925]
[926]
[927]
[928]
[929]
[930]
[931]
[932]
[933]
[934]
[935]
[936]
[937]
[938]
[939]
[940]
[941]
[942]
[943]
[944]
[945]
[946]
[947]
[948]
[949]
[950]
[951]
[952]
[953]
[954]
[955]
[956]
[957]
[958]
[959]
[960]
[961]
[962]
[963]
[964]
[965]
[966]
[967]
[968]
[969]
[970]
[971]
[972]
[973]
[974]
[975]
[976]
[977]
[978]
[979]
[980]
[981]
[982]
[983]
[984]
[985]
[986]
[987]
[988]
[989]
[990]
[991]
[992]
[993]
[994]
[995]
[996]
[997]
[998]
[999]
[1000]

Qy	593	EEAI---TPOASBEL--ASMELNESSRWTE--EMETAKGLLEHGRNWSAIAIRMGSKTV	646
Db	763	DEAVHKASQOAKRLPPAISRKSMTSLEEMVSPWKKBIEYEQPAAVSPMEKRTV	822
Qy	647	SQKNFYFNKYQRONDEILLQOHLKMEKERNARKKKAPAAASEBAZAPPVVUDEEME	706
Db	823	YQ-----MALNKLDEILL-----AAQQTIS-----ASE	845
Qy	707	ASGVSGNEEMVEAEALHASGNEVPREGCSGPATVNNSSDTSISPHTAAAKDTQONG	766
Db	846	SPGPGG-----LASLGRKRPKGFFA-----TSSPDPHRS-----QPS	879
Qy	767	PKPATIGADGPPPGPTPPRRTS-----RAPIETPPA---SEATGAPTPPPAPSPSPA	817
Db	880	YDRSFL-----PPGGLMLRQKSIGAAEDDRPYLAPPAMKFSRSLSVPGSEIDIPPPPT	934
Qy	818	PPVVVPEEKEEETAAPAPPVEEGEQPPAAEELAVDTGKAEPVSEKTEERAEEGPAKG	877
Db	935	SPP-----BPPY-----STPPA-----	946
Qy	878	KDAEAEATAEGALKAEKGGGRATTA-----KSGAPODSDSSATCSADEVDEAE	930
Db	947	-----PSSSRLTPSRGGFPNPSGGPLPASPSFSDGSPDPDR	987
Qy	931	GGDKNRLSPRSLTTPGDFRANASQKPLDLKQLKQRAAAIPIPIQTVKVHPP	987
Db	988	GGGREKSL-----YHSAALPPAH-----HHPPHHHH	1014
Qy	988	DAAPTGP-----APPAPPPQNLQEPESDAPQPGSGPKGRSRPAPPA	1030
Db	1015	HAPPQPHHHHAHPHPPEMTGSGDDPRLALGPQSLRWGSGPSTSGAPSPSH	1074
Qy	1031	DKEAFAEAQKLRGDPDCWTSGLFPFVPREVIKASHPADPGAFSAVP	1082
Db	1075	HSSGGSG---PTQAPALRY---FQLPPR-ASAAMVVPARGGRKKGLPVQTKVEGE	1127
Qy	1083	P-----LPLGLHTDARVLPPTTISNPPLISSAKH-----PSVLER-----QIGAISQGM	1129
Db	1128	PQKGSIFSASSPTS-PALPR---SEPPAGPSEKNSIPIPTIIKAPSTSSGRSSQGS	1182
Qy	1130	SVQLHVYSEH-----AKAPGVVTWGLPLPMDPKLAPFGVKQEQLSPRG	1176
Db	1183	STEAEPPTQDAGAGGSGSPAPATSPVPSPSPVPTPASPGPATLDTSFQGAALVG	1242
Qy	1177	QAGPESIGVPTAQEAASVLRGALSGVPGGSITKGIPSTRVPSDSAITYRGSITHGTPAD	1236
Db	1243	AARREGWQNEARRRSTLFTLSTDAGDBDGDGSLG-FCG--PPGPRUHRKSIDEG----	1295
Qy	1237	VLYKGTITRIIGEDSPRLDGRBDSLPGKHVIYEGKHGHLVSGWMSVTCQSKEDGRS	1296
Db	1296	-----MFAEPLYRLSESG-----GSSGGYGAYAAG-----SRAYGGS	1327
Qy	1297	SSGPPHETAAPKRTYDMWGRVGRBIAISSIEGLM-----GRAIPPER--	1339
Db	1328	GSSSAFTSFLPPR--PLVHPLTGLKALDPASPLGLAARERALKESSEGCGTQPpprpp	1385
Qy	1340	-----HSPH--HLKEQHHR-----GSITQIGIPRSYVE	1365
Db	1386	SPRYDAPPTLHHSHSPHSPHARHFPVLRMLMGDPRRELVGRVAGLGSQEKALTASPPA	1445
Qy	1366	AQEDYLRR-----EAKLKREGTPPPPPSRDLTEAYKTOALGPLKLPAHEGLVATV	1418
Db	1446	ARRSLHLRLPTAPGVGFLLLQLG-PEPPTPHGVSKAWRTAA-----	1487
Qy	1419	KEAGRSIHEIPREELRHTPE-LPLAPRLKEGSIQTGTPLKYDTGASTTGSKKHVRSLI	1477
Db	1488	-----PEE-----PERLPLHVRFLEN---CQARP---PPAGTRGSSTED----	1520
Qy	1478	GSPGRTPPVPVHPLDVMADARALERACVYESLSKRPPTASSGGSGIARGAPVIVPELGKPR	1537
Db	1521	-GPG--VPPSPRRVL-----PTSPTSFRGNEENGFLLV-----LPP	1555

RESULT 20

RESOUL 20
MLL4 HUMAN

	MLL4_HUMAN	PRT; 2715 AA.
ID_	MLL4 HUMAN	STANDARD;

Q9UMN6; O15022; Q95836; Q96GP2; Q96IP3; Q9UK25; Q9Y668; Q9Y669;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Myeloid/lymphoid or mixed-lineage leukemia protein 4 (Trithorax

DE homology 2).

GN MLL4 OR TRX2 OR HRX2 OR MLL2 OR KIAA0304.

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OS Homo sapiens (human).

Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Hominidae; Homo.

OC Mammalia, Eutheria
OX NCBI TaxID=9606;

OX NCBI-
RN [1]

RN [T]
RP SEQUENCE FROM N.A. (ISOFORM 1):

RA Angrand P.O., Valvatne H., Jeanmougin F., Adamson A.,
RA van der Hoeven F., Olsen L., Tekotte H., Huang N., Poch O.,
RA Lamerdin J., Chambon P., Lossion R., Stewart A., Aasland R.,
RT "Mammalian trithorax- and ASH1-like proteins: putative chromatin
RT regulators which contain PHD fingers and SET domains";
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RA Lamerdin J.E., McCready P.M., Adamson A.W., Burkhardt-Schultz K.,
RA Garcia E., Kyle A., Ramirez M., Stilwagen S., Garnes J., Danganan L.,
RA Bruce R., Quan G., Montgomery M., Ow D., Kobayashi A., Olsen A.O.,
RA Carrano A.V.;
RT "Sequence analysis of a 1 Mb region in human 19q13.1";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 111-2715 FROM N.A. (ISOFORM 1).
RC TISSUE=Leukocyte, and Testis;
RX MEDLINE=20105772; PubMed=10637508;
RA Huntsman D.G., Chin S.-F., Mulleris M., Batley S.J., Collins V.P.,
RA Wiedemann L.M., Aparicio S., Caldas C.;
RT "MLL2, the second human homolog of the Drosophila trithorax gene, maps
RT to 19q13.1 and is amplified in solid tumor cell lines";
RL Oncogene 18:7975-7984(1999).
RN [4]
RP SEQUENCE OF 816-2715 FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=97349984; PubMed=9205941;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
RT The complete sequences of 100 new cDNA clones from brain which can
RT code for large proteins in vitro";
RL DNA Res. 4:141-150(1997).
RN [5]
RP SEQUENCE OF 1918-2715 FROM N.A.
RC TISSUE=Brain, and Skin;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins P.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Roderiguez A.C., Grinstead J., Schmutz J., Myers R.M.,
RA Buterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [6]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Bone marrow, and Placenta;
RX MEDLINE=99339983; PubMed=10409430;
RA FitzGerald K.T., Diaz M.O.;
RT "MLL2: A new mammalian member of the trx/MLL family of genes";
RL Genomics 59:187-192(1999).
CC -1- FUNCTION: Possibly acts as a transcriptional regulatory factor.
CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=1; Synonyms=Long;
CC IsoId=O9UNM6-1; Sequence=Displayed;
CC Name=2; Synonyms=Truncated;
CC IsoId=O9UNM6-2; Sequence=VSP_006668, VSP_006669;
CC -1- TISSUE SPECIFICITY: Widely expressed. Highest levels in testis.

CC Also found in brain, bone marrow, heart, muscle, kidney, pancreas,
CC spleen, thymus, prostate, ovary, intestine, colon, peripheral
CC blood lymphocytes, and placenta.
CC -1- DISEASE: Often amplified in pancreatic carcinomas.
CC -1- SIMILARITY: Belongs to the TRX/MLL family.
CC -1- SIMILARITY: Contains 1 bromodomain.
CC -1- SIMILARITY: Contains 1 PHD-type zinc fingers.
CC -1- SIMILARITY: Contains 3 PHD-type zinc fingers.
CC -1- SIMILARITY: Contains 1 CXC-type zinc finger.
CC -1- SIMILARITY: Contains 1 post-SET domain.
CC -1- CAUTION: This protein was first named MLL2 by Ref.3 and Ref.6.
CC MLL2 corresponds to another protein located on chromosome 12 (see
CC AC 014686).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AJ007041; CAB45385.1; -;
CC EMBL; AD000671; -; NOT_ANNOTATED_CDS.
CC EMBL; AB002302; BAA20763.2; -;
CC EMBL; AF186605; AAD56420.1; -;
CC EMBL; AF104918; AAD17932.1; -;
CC EMBL; AF105279; AAD26113.1; -;
CC EMBL; BC009337; AAH09337.1; -;
CC EMBL; BC007353; AAH07353.1; -;
CC EMBL; AF105280; AAD26112.1; -;
CC MIM; 606834; -;
CC GO; GO:0005634; C:nucleus; NAS.
CC GO; GO:0003700; F:transcription factor activity; NAS.
CC GO; GO:0008270; F:zinc ion binding; NAS.
CC GO; GO:0048096; P:chromatin-mediated maintenance of transcript. . .; NAS.
CC InterPro; IPR003889; FyricH.C.
CC InterPro; IPR003888; FyricH.N.
CC InterPro; IPR003616; PostSET.
CC InterPro; IPR001214; SET.
CC InterPro; IPR002857; Znf_CXHC.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF00628; PHD; 3.
CC Pfam; PF00856; SET; 1.
CC Pfam; PF02008; zf-CXHC; 1.
CC SMART; SM00542; FYRC; 1.
CC SMART; SM00541; FYRN; 1.
CC SMART; SM00249; PHD; 4.
CC SMART; SM00508; PostSET; 1.
CC SMART; SM00317; SET; 1.
CC PROSITE; PS00868; POST_SET; 1.
CC PROSITE; PS0280; SET_1.
CC PROSITE; PS01359; ZF_PHD_1; 3.
CC PROSITE; PS00016; ZF_PHD_2; 3.
CC DNA-binding; Bromodomain; Nuclear protein; Zinc-finger; Metal-binding;
CC Transcription regulation; Alternative splicing; Repeat.
CC A.T HOOK (BY SIMILARITY).
CC DNA_BIND 37 44
CC 110 117
CC 357 365
CC DNAS_BIND 357 365
CC DNAS_BIND 959 1005
CC ZN_FING 1201 1252
CC ZN_FING 1249 1303
CC ZN_FING 1335 1396
CC ZN_FING 1449 1471
CC DOMAIN 2574 2695
CC DOMAIN 2699 2715
CC DOMAIN 26 37
CC DOMAIN 248 255
CC DOMAIN 362 398
CC DOMAIN 402 771
CC DOMAIN 808 812
CC DOMAIN 1963 1970
CC DOMAIN 2251 2259
CC POLY-GLY.
CC POLY-PRO.
CC ASP/GLU-RICH (ACIDIC).
CC PRO-RICH.
CC POLY-GLN.
CC POLY-PRO.
CC POLY-PRO.

Db 1782 EYRWGPF---REEPAHLEAAEENQTIIVHSPAPSEPPGG----- 1817
Qy 1803 ERDRREKESILTSITVEHAPLWR---PGTEQSSGSSGGGGSSRRSPASHAHQH 1860
Db 1818 ---EDPPLDLDVLPVGPAPERHSPIONIDPLLRPDGSA----- 1852
Qy 1861 SPISPRDQALQRPSPVLHNTGMGIITAVEPSPKTVLRLSTSTSPVRPAATPPATHCP 1920
Db 1953 PPPAPRS-----FSGARIKVPNTSPSR-----P 1876
Qy 1921 LGGTLGDVYPTLMPVLLPKAPRVARPERPRADTGHAFIAKPPAR-----SGLEPASPPS 1976
Db 1877 LGGVSFGPLPGSPFSSLTTHIPTVGDPDF-----APRRGRRPSPLAPRPPS 1926
Qy 1977 KGSEPRLL-----VPPVSGHATITPAKNLAPHASDPDPAP----- 2014
Db 1927 RWASP-FLKTSPPQURVPPPTSVVLTFTSGELAPPGPAPSPPPEDLGDDEFMEVVG 1985
Qy 2015 -PASASDPHREKTKOSKPFISQELRLSLGHGSSYSPEGVE-----PVSPV 2059
Db 1986 LSAADLFAASLLGTEFF--QE-EIVAAGAMGSSHGSGGDSSEESSPTSRYIHFPVTV 2042
Qy 2060 SSPSL--THDKGLPKHLELD-----KSHLEGELPK-----QPGPVKLGEAAHLP 2104
Db 2043 SAPGLAPSATGAPR-IEQLDGVDDGTDEAEAVQPRGQGTTPSPGPGVVRAGVLGAAGD 2101
Qy 2105 HLRPLPESQSSPLLQTPAGVKHQVRVTLAQHISEVITDYTEHHHPQOLSAPLPAPLY 2164
Db 2102 RARP-PEDLP-----SEIV--DFVL-----KNLGGP----- 2124
Qy 2165 SFGASCPCVLDLRPPPSDLYLPDPDHGAP--ARGSPHSEGGKSP-BPNKTSVLGGGEGD 2221
Db 2125 GDGGAG-----PREESLPP-----APPLANGSQPSQGLTASPADPTRTFAMLPAGP 2171
Qy 2222 IEPVS-----PPGEMTEP--GHSRAVYPLLVRDGEQTPSRMGSKSPGNTSQPPAF 2272
Db 2172 VRVLSLGPAPPPKPAATSKILVNKLQGVFVMAAGEGEPVPPP-----VKQPLP-- 2220
Qy 2273 SKLTESAMVKKQEKINKLNTNHNREPNYISQPTTEFNMPAITGTGLMYRSQAV 2332
Db 2221 -----PPTISPTAPTSWTLPGPLGLVPLV----- 2246
Qy 2333 QEHASTNMLBAIRKA-----LMGKYDQWEEESPPLSA-----NAFNPLNASALPA 2379
Db 2247 -----GVVRPAPPPPPPPPLTLVLSSGSPASPPRQAIRVKRVSTFSKSPAPPPY 2295
Qy 2380 AMPITAADGR--SDHTLTSFGGGKAKVSCRPSRRKAKSP-APGLASGDRPPSVSVHSEG 2437
Db 2296 KAPRLDEGDEASEDTPQVPEGLG-----SGGFSRVRMKPTTVRGVLDLDRPGEPAGEESPG 2350
Qy 2438 DCNRRTPLTNRVWEDRRSSAGSTPPYPNPLIMRLQ-----AGVMAS-----PPPPGLPAGSG 2489
Db 2351 PLQERSPLL-PLPEDGPPQVDPGP---PDLLLESQWHYSGEASSSEEPSPDDKENQA 2406
Qy 2490 P-LAGPH 2495
Db 2407 PKRTGPH 2413

RESULT 21

NCO6 HUMAN

ID NCO6 HUMAN STANDARD; PRT: 2063 AA.

AC Q14686; Q9UNT9; Q9UH74; Q9UK86;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 15-MAR-2004 (Rel. 43, Last annotation update)

DE Nuclear receptor coactivator 6 (Amplified in breast cancer-3 protein)

DE (Cancer-amplified transcriptional coactivator ASC-2) (Activating

DE signal cointegrator-2) (ASC-2) (Peroxisome proliferator-activated

DE receptor-interacting protein) (PPAR-interacting protein) (PRIP)

DE (Nuclear receptor-activating protein, 250 kDa) (Nuclear receptor

DE coactivator RAP250) (NRC RAP250) (Thyroid hormone receptor-binding

DE protein).

GN NCOA6 OR AIB3 OR RAP250 OR TRBP OR KIAA0181.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND INTERACTION WITH CREBBP; NCOA1; GTF2A; TBP;
RP RXRA; ESR1; RARA AND THRA.
RX MEDLINE=20036574; PubMed=10567404;
RA Lee S.-K., Anzick S.L., Choi J.-E., Bubendorf L., Guan X.-Y.,
Jung Y.-X., Kallioniemi O.P., Kononen J., Trent J.M., Azorsa D.,
Jhun B.-H., Cheong J.H., Lee Y.C., Meltzer P.S., Lee J.W.;
RT "A nuclear factor ASC-2, as a cancer-amplified transcriptional
RT coactivator essential for ligand-dependent transactivation by nuclear
RT receptors in vivo";
RL J. Biol. Chem. 274:34283-34293 (1999).
RN [2]
RP SEQUENCE FROM N.A., HOMODIMERIZATION, AND INTERACTION WITH CREBBP;
RP RXRA; ESR1; NR3C1; RARA; VDR AND THRA.
RX MEDLINE=200325329; PubMed=10866662;
RA Mahajan M.A., Samuels H.H.;
RT "A new family of nuclear receptor coregulators that integrates nuclear
RT receptor signaling through CBP";
RL Mol. Cell. Biol. 20:5048-5063 (2000).
RN [3]
RP SEQUENCE FROM N.A., AND INTERACTION WITH PPARA; PPARG; ESR1; ESR2 AND
RP THRA.
RX TISSUE=Testis;
RX MEDLINE=20148724; PubMed=10681503;
RA Cairns F., Antonson P., Pelto-Huikko M., Treuter E., Gustafsson J.-A.;
RT "Cloning and characterization of RAP250, a nuclear receptor
RT coactivator";
RL J. Biol. Chem. 275:5308-5317 (2000).
RN [4]
RP SEQUENCE FROM N.A., PHOSPHORYLATION BY PRKDC, AND INTERACTION WITH
RP THR; RAR; EP300 AND CRSP3.
RX TISSUE=Lymphocytes;
RX MEDLINE=20283976; PubMed=10823961;
RA Ko L., Cardona G.R., Chin W.W.;
RT "Thyroid hormone receptor-binding protein, an LXXLL motif-containing
RT protein, functions as a general coactivator";
RL Proc. Natl. Acad. Sci. U.S.A. 97:6212-6217 (2000).
RN [5]
RP SEQUENCE FROM N.A.
RP TISSUE=Bone marrow;
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1";
RL DNA Res. 3:17-24 (1996).
RN [6]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.B.,
Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
Lehvasalo M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
Marsh V.L., Martin S.L., McConachie L.J., McClay K., McMurray A.A.,
Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
Olliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Rameay H.,
Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,

Db 1356 -----VSEFSEAKDENASISPMDEBVPDSEFVE-KVLSPLRSPFLLGSESPY-- 1405
 Qy 1366 AQEDYLREAKLKR-----EGTTPPPPSRLDTEAYKTOALGPLKLPKPAHEG 1413
 Db 1406 --EDFLSADSKVLGRRESPEFGKNGKQGFDPRESVSDUT-----STGLYQDKQEKSTG 1459
 Qy 1414 LVATVKEAGRSIHEIPREELRHTEPELAPRLPEKGSITGCTPLKYDTGASTTGSKKHDV 1473
 Db 1460 FIPKEDFGPEKTKSDVETWSSQSALADERKL-GGDV-----SPTQID--VSQFGSPKEDT 1513
 Qy 1474 RSLIGSPQRTPPVPHPLVDNADAPALERACVRESLKRPTGASSGSGSIARGAPVIVPEL 1533
 Db 1514 KWSISEGTVSKSATPD-----EGVAEDTSHMEGVASVSTASVATSS---FPPEP 1561
 Qy 1534 GKQPSPLTYDHGAPAGHLPGRSPVTMEPTPRLOEGLSSSKASQDKRLSTPREI- 1592
 Db 1562 TTDDVPSLHAEGVSPHSTEVDDSLSVSVVQTPPTFQETEMSPKCECPRMISIPDPS 1621
 Qy 1593 --AKSPHSTVPEHHPHPISPYEHLLRGVGDVLYRSHIPLAFDPTSPRGIPLDAAAY 1649
 Db 1622 PKTAKS-RTPVDHRSE-----QSSMSIEFGQESPESHFAMDFS--- 1659
 Qy 1650 YLPHLANPYPHYLPYLRGVPDPALENQTIINDYITSQMHNTATAMAQADM 1709
 Db 1660 -----RQSDPHPTL-----GASVLHITENGTEV-DY----- 1685
 Qy 1710 LRGLSP---RSSIALNVAAGRGIIIDLSQVPHLPVLPVPPPGTPATAMDRLAYLTPAQ 1766
 Db 1686 ----SPCDIQDSSLHK-----IPPEEPTSYIQDNDLSLISVSQ 1721
 Qy 1767 ----PFSRHS-----SSPLSPGGGTHLTKPTTS-----SSERERDRDRDREREK 1812
 Db 1722 VEASPTSSAHTPSQIAPLOBDTLSDVVPREMSLYASLASEKVSQLEGE-----KLSPK 1777
 Qy 1813 SILTSTTVEHAPIWRPTEQSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 1872
 Db 1778 SDISPLTPRESSPLSPGSDSTAETAAMAHQASSP-----PIDATAEPYG 1827
 Qy 1873 QRPVNLHNTGMGIITAVEPSKPTVLRSTSTSPVRPAATPPATHCPGLGTLGCVYPTL 1932
 Db 1828 FRSSMLFDTWQHL-----ALNRDLTSSVEKSGGKTP-----GDFNYAYQKP 1871
 Qy 1933 MEVLLPKBA-----PRVAPERPRADTGHAF-----LAKPPAR 1966
 Db 1872 ENAAGSPDEEDYDSEKTIIRHDVVRVYVEKTERTKSPCDSGSYHETIKTKTPE 1931
 Qy 1967 SGL-----EPASSPSKG-----SEPRPLVPPVSGHATARTAPKNLAPH----- 2005
 Db 1932 GGYTCITETKTRTPPEGYSVEISEKTRTPVSGYTYEKTERRLLDDISNGYDDTE 1991
 Qy 2006 ---HASPDPPAPPASADPHREKTSQKPSIQEILRSLGYH-----GSSYSPEG 2052
 Db 1992 DGGHTLGD-----CSYSYETEKITSFP-----ESESYSYETSTKTRSPTSAYCYET 2040
 Qy 2053 VEPVSPVSPS-----LTHDKGLPKH-----LBELDKSHLEGLRPK--OP 2091
 Db 2041 MEKITTPQASTVSYETSDRYTEKTSSEARQVDVLCVSCFEKPKTELSPSFNP 2100
 Qy 2092 GPVK-LGGEAAHLPLRPLPESQPSPLQOT--APGVKHQRVWVLAQHISEVITQDYT 2148
 Db 2101 NPLEWPAGE-----EPTSESE---KPLTQSGGAPPPSGGK---QGRQCDTEPTTSVS 2147
 Qy 2149 RHHPQULSALPAPLYSPFGASCPLV-----DLRRPPSDLYL-----PPPDHGA 2192
 Db 2148 ESAPSQSDVPPPT-----EECPSITADANIDSESESETIPTDKTVYKHMDDPP--A 2199
 Qy 2193 PARGPSHSEGGKSPENKTSVLGGGBDGI-----EPVSPPEGMTEPCHERSAV 2241
 Db 2200 PMQ-----DRSPSRHPDVSVMVDPDALAVQNLGKAVKDKLKEKTKTKGPTKSS 2251
 Qy 2242 YPLLRYGCEQTEPBRMGSKSPGNTSQPPAFPSKLTESNAWKSQKQKINKLNTHRNE 2301

Db 2252 SPVKKGDGK-----SKFLAASPFGALKESSDKSVRSVPSPKSKESVEKA-TKTTTT 2301
 Qy 2302 PEYNIQSGTETFMNPALTGTGLMTYRSQAOEHASTNMGLEAIRKALMGKYDOWEESP 2361
 Db 2302 PEV-----KATRG-----BEKD 2313
 Qy 2362 PLSANAENPLNASLASLPAAMBITAAGRSDHTLTPSGGGGKAKVGRPSSSKAKSPAPGL 2421
 Db 2314 KETQNA-----ANASAKSKAK--TAT-----TGPGTKTKAKSTVP-----PGL 2350

RESULT 23
 SNT1 YEAST
 ID SNT1 YEAST STANDARD; PRT: 1226 AA.
 AC P25357; Q02397; Q8N1L8;
 DT 10-MAY-1992 (Rel. 22, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DE Potential DNA-binding protein SNT1.
 GN SNT1 OR YCR033W OR YCR33W OR YCR592.
 OS Saccharomyces cerevisiae (Baker's Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92133166; PubMed=1776366;
 RA Wicksteed B.L., Roberts A.B., Sagliocco F.A., Brown A.J.P.;
 RT "The complete sequence of a 7.5 kb region of chromosome III from
 Saccharomyces cerevisiae that lies between CRY1 and MAT.";
 RL Yeast 7:761-772(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91335897; PubMed=1872032;
 RA Jia Y., Slonimski P.P., Herbert C.J.;
 RT "The complete sequence of the unit YCR59, situated between CRY1 and
 MAT, reveals two long open reading frames, which cover 91% of the
 10.1 kb segment.";
 RL Yeast 7:413-424(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C;
 RX MEDLINE=92244356; PubMed=1574125;
 RA Oliver S.G., van der Aart Q.J.M., Agostoni-Carbone M.L., Aigle M.,
 RA Alberghina L., Alexandraki D., Antoine G., Anwar R., Ballesta J.P.G.,
 RA Benit P., Berben G., Bergantino E., Biteau N., Bolle P.-A.,
 RA Bolotin-Fukuhara M., Brown A.J.P., Brown R., Buhler J.-M.,
 RA Carignani G., Chanet R., Contreras R., Crouzet M., Daignan-Fornier B.,
 RA De Haan M., Defoor E., Delgado M.D., Demolder J., Doira C., Dubois E.,
 RA Dujon B., Duesterhoeft A., Erdmann D., Fiers W.,
 RA Fairhead C.A., Faye G., Feldmann H., Frontali L., Fukuhara H.,
 RA Francinguet-Gaillard M.-C., Franco L., Giliquet V., Glandsdorff N.,
 RA Fuller L.J., Gent M.E., Gigot D., Gilliquet V., Grivell L.A., Haesemann M.,
 RA Goffeau A., Gresson M., Grisanti P., Gruber C.J., Hilger F., Hohmann S.,
 RA Hatat D., Hegemann J.H., Herbert C.J., Indje K.J., Isono K., Jackman P.,
 RA Hollenberg C.F., Huse K., Iborra F., Jauniaux J.-C., Jia Y., Jimenez A.,
 RA Jaq C., Jacquet M., James C.M., Jauniaux J.-C., Lewis C., van der Linden C.G.,
 RA Kleinhaus U., Kreisl P., Lafranchi G., Lewis C.,
 RA Lucchini G., Lutkenkirchen K., Maat C.C., Mannheim G., Manzano M.E.,
 RA Martegani E., Mathieu A., Maurer C.T.C., McConnell D., McKee R.A.,
 RA Mesenguy P., Mewes H.-W., Molemans F., Montague M.A., Navas L.,
 RA Newlon C.S., Olson M.V., Pallier C., Panzeri L., Pearson B.M.,
 RA Perea J., Philippsen P., Pierard A., Planta R.J., Plevani P.,
 RA Poetsch B., Pohl F.M., Purnelle B., Ramezani Rad M., Rasmussen S.W.,
 RA Raynal A., Remacha M., Richterich P., Roberts A.B., Rodriguez F.,
 RA Sanz E., Schaaff-Gerstenschlaeger I., Scherens B., Schweitzer B.,
 RA Shu Y., Skala J., Slonimski P.P., Sor F., Soustelle C.,
 RA Spiegelsberg R., Stave L.I., Steensma H.Y., Steiner S., Thierry A.,
 RA Thireos G., Triano L.N., Urrestazu L.A., Valle G., Vetter I.,
 RA van Vliet-Needijk J.C., Volckaert G., Vreken P., Warrington J.R.,
 RA von Wettstein D., Wicksteed B.L., Wilson C., Wurst H., Xu G.,
 RA Zimmermann F.K., Sgouros J.G.;
 RT "The complete DNA sequence of yeast chromosome III."

RESULT 26	PCLO_CHICK	STANDARD;	PRT;	5120 AA.	FT DOMAIN	4414	4493	PDZ.
ID	PCLO_CHICK				FT DOMAIN	4627	4726	C2 DOMAIN 1.
AC	Q9PU36;				FT DOMAIN	5003	5094	C2 DOMAIN 2.
DT	28-FEB-2003 (Rel. 41, Created)				SQ SEQUENCE	5120	AA;	560751 MW; A658D9891B65B412 CRC64;
DT	28-FEB-2003 (Rel. 41, Last sequence update)							
DT	28-FEB-2003 (Rel. 41, Last annotation update)							
DE	Piccolo protein (Acronin) (Fragment).							
DE	PCLO OR ACZ.							
GN	Gallus gallus (Chicken).							
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;							
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;							
OC	Gallus.							
OX	NCBI TaxID=9031;							
RN	[1]							
RP	SEQUENCE FROM N.A.							
RC	TISSUE=Brain;							
RX	MEDLINE=99439764; PubMed=10508862;							
RA	Wang X., Kibschull M., Laue M.M., Lichte B., Petrasch-Parwez E.,							
RA	Kilimann M.W.;							
RT	"Acronin, a 550-kd putative scaffolding protein of presynaptic active							
RT	zones, shares homology regions with rim and bassoon and binds							
RT	profilin.";							
RL	J. Cell Biol. 147:151-162(1999).							
CC	FUNCTION: May act as a scaffolding protein involved in the							
CC	organization of synaptic active zones and in synaptic vesicle							
CC	trafficking (By similarity).							
CC	SUBUNIT: Interacts with Rabac1/Pral and profilin (By similarity).							
CC	SUBCELLULAR LOCATION: Concentrated at the presynaptic side of							
CC	synaptic junctions (By similarity).							
CC	DOMAIN: C2 domain 1 is involved in binding calcium and							
CC	phospholipids. Calcium binds with low affinity but with high							
CC	specificity and induces a large conformational change.							
CC	SIMILARITY: Contains 2 C2 domains.							
CC	SIMILARITY: Contains 1 PDZ/DRH domain.							
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration							
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -							
CC	the European Bioinformatics Institute. There are no restrictions on its							
CC	use by non-profit institutions as long as its content is in no way							
CC	modified and this statement is not removed. Usage by and for commercial							
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/							
CC	or send an email to license@isb-sib.ch).							
CC	EMBL; Y19187; CAB60725.1; ..							
DR	HSP; P04410; 1A25.							
DR	GO; GO:0045202; C:synaptic junction; ISS.							
DR	GO; GO:0005509; F:calcium ion binding; ISS.							
DR	GO; GO:0005544; F:calcium-dependent phospholipid binding; ISS.							
DR	GO; GO:0005522; F:profilin binding; ISS.							
DR	GO; GO:0007010; P:cytoskeleton organization and biogenesis; ISS.							
DR	GO; GO:0016080; P:synaptic vesicle targeting; ISS.							
DR	InterPro; IPR000008; C2.							
DR	InterPro; IPR001478; PDZ.							
DR	InterPro; IPR001565; Synaptotagmin.							
DR	InterPro; IPR008899; Znf_piccolo.							
DR	Pfam; PF00168; C2; 2.							
DR	Pfam; PF00595; PDZ; 1.							
DR	Pfam; PF05715; Zf_piccolo; 2.							
DR	PRINTS; PR00399; SYNAPTOTAGMN.							
DR	SMART; SM00239; C2; 2.							
DR	SMART; SM00228; PDZ; 1.							
DR	PROSITE; PS00499; C2 DOMAIN 1; 1.							
DR	PROSITE; PS00004; C2 DOMAIN_2; 2.							
DR	PROSITE; PS01005; PDZ; 1.							
KW	Calcium/phospholipid-binding; Zinc; Metal-binding; Zinc-finger;							
KW	Repeat.							
FT	NON TER							
FT	DOMAIN	1	357	10 X 10 AA TANDEM APPROXIMATE REPEATS OF				
FT		258		P-A-K-P-Q-P-Q-P-X.				
FT	ZN_FING	368	392	C4-TYPE (POTENTIAL).				
FT	ZN_FING	836	859	C4-TYPE (POTENTIAL).				
FT	DOMAIN	2324	2343	POLY-PRO.				

DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE OF 1-472 FROM N.A.
RX MEDLINE=89025644; PubMed=3178743;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RA Jaenisch R., Prockop D.J.;
RT "Structure of a full-length cDNA clone for the prepro alpha 1(I)
RT chain of human type I procollagen.";
RL Biochem. J. 253:919-922(1988).
RN [2]
RP SEQUENCE OF 1-181 FROM N.A.
RX MEDLINE=84270697; PubMed=6462220;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ding J.-F., Morabito M.,
RA Myers J., Williams C., Ramirez F.;
RT "Human pro alpha 1(I) collagen gene structure reveals evolutionary
RT conservation of a pattern of introns and exons.";
RL Nature 310:337-340(1984).
RN [3]
RP SEQUENCE OF 162-301.
RC TISSUE=Skin;
RX MEDLINE=71038625; PubMed=5529814;
RA "Isolation and characterization of the cyanogen bromide peptides from
RT the alpha 1 and alpha 2 chains of human skin collagen.";
RL Biochemistry 9:4699-4706(1970).
RN [4]
RP SEQUENCE OF 263-268.
RC TISSUE=Skin;
RX MEDLINE=71001508; PubMed=4319110;
RA Morgan P.H., Jacobs H.G., Segrest J.P., Cunningham L.W.;
RT "A comparative study of glycopeptides derived from selected
RT vertebrate collagens. A possible role of the carbohydrate in fibril
RT formation.";
RL J. Biol. Chem. 245:5042-5048(1970).
RN [5]
RP SEQUENCE OF 425-1464 FROM N.A.
RX MEDLINE=84080385; PubMed=6689127;
RA Bernard M.P., Chu M.-L., Myers J.C., Ramirez F., Eikenberry E.F.,
RA Prockop D.J.;
RT "Nucleotide sequences of complementary deoxyribonucleic acids for the
RT pro alpha 1 chain of human type I procollagen. Statistical evaluation
RT of structures that are conserved during evolution.";
RL Biochemistry 22:5213-5223(1983).
RN [6]
RP SEQUENCE OF 1229-1454 FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=88124208; PubMed=3340531;
RA Mäkelä J.K., Raassina M., Virta A., Vuorio E.;
RT "Human pro alpha 1(I) collagen: cDNA sequence for the C-propeptide
RT domain.";
RL Nucleic Acids Res. 16:349-349(1988).
RN [7]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88097389; PubMed=3480516;
RA Bornstein P., McKay J., Morishima J.K., Devayayalu S., Gelinak R.E.;
RT "Regulatory elements in the first intron contribute to
RT transcriptional control of the human alpha 1(I) collagen gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:8869-8873(1987).
RN [8]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=85130970; PubMed=2857713;
RA Chu M.-L., de Wet W.J., Bernard M.P., Ramirez F.;
RT "Fine structural analysis of the human pro-alpha 1 (I) collagen gene.
RT Promoter structure, Alu repeats, and polymorphic transcripts.";
RL J. Biol. Chem. 260:2315-2320(1985).
RN [9]
RP SEQUENCE OF 1-44 FROM N.A.
RX MEDLINE=88033098; PubMed=2822714;
RA Rossouw C.M.S., Vergeer W.P., du Plooy S.J., Bernard M.P., Ramirez F.,
RA de Wet W.J.;
RT "DNA sequences in the first intron of the human pro-alpha 1(I)
RT collagen gene enhance transcription.";
RL J. Biol. Chem. 262:15151-15157(1987).
RN [10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
RN [11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [12]
RP REVIEW ON VARIANTS.
RX MEDLINE=91374476; PubMed=1895312;
RA Byers P.H., Wallis G.A., Willing M.C.;
RT "Osteogenesis imperfecta: translation of mutation to phenotype.";
RL J. Med. Genet. 28:433-442(1991).
RN [13]
RP REVIEW ON VARIANTS.
RX MEDLINE=97169389; PubMed=9016532;
RA Dalglish R.;
RT "The human type I collagen mutation database.";
RL Nucleic Acids Res. 25:181-187(1997).
RN [14]
RP VARIANT OI-II CYS-1166.
RX MEDLINE=86287390; PubMed=3016737;
RA Cohn D.H., Byers P.H., Steinmann B., Gelinak R.E.;
RT "Lethal osteogenesis imperfecta resulting from a single nucleotide
RT change in one human pro alpha 1(I) collagen allele.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:6045-6047(1986).
RN [15]
RP VARIANT OI-II ARG-569.
RX MEDLINE=87222295; PubMed=3108247;
RA Bateman J.F., Chan D., Walkers J.G., Cole W.G.;
RT "Lethal perinatal osteogenesis imperfecta due to the substitution of
RT arginine for glycine at residue 391 of the alpha 1(I) chain of type I
RT collagen.";
RL J. Biol. Chem. 262:7021-7027(1987).
RN [16]
RP VARIANT OI-II CYS-926.
RX MEDLINE=88033031; PubMed=3667599;
RA Vogel B.E., Minor R.R., Freund M., Prockop D.J.;
RT "A point mutation in a type I procollagen gene converts glycine 748
RT of the alpha 1 chain to cysteine and destabilizes the triple helix in
RT a lethal variant of osteogenesis imperfecta.";
RL J. Biol. Chem. 262:14737-14744(1987).
RN [17]
RP VARIANT OI-II ARG-842.
RX MEDLINE=88298828; PubMed=3403550;
RA Bateman J.F., Lemande S.R., Dahl H.H., Chan D., Cole W.G.;
RT "Substitution of arginine for glycine 664 in the collagen alpha 1(I)
RT chain in lethal perinatal osteogenesis imperfecta. Demonstration of
RT the peptide defect by in vitro expression of the mutant cDNA.";
RL J. Biol. Chem. 263:11627-11630(1988).
RN [18]
RP VARIANT OI CYS-1195.
RX MEDLINE=89218628; PubMed=3244312;
RA Labhard M.E., Wirtz M.K., Pope F.M., Nicholls A.C., Hollister D.W.;
RT "A cysteine for glycine substitution at position 1017 in an alpha
RT 1(I) chain of type I collagen in a patient with mild dominantly
RT inherited osteogenesis imperfecta.";

MLL3_MOUSE	STANDARD; PRT; 4903 AA.	CC
AC	QBRH4; QBRK12; Q8G6M3; Q923H5; Q923H6;	CC
DT	10-OCT-2003 (Rel. 42, Created)	CC
DT	10-OCT-2003 (Rel. 42, Last sequence update)	CC
DT	15-MAR-2004 (Rel. 43, Last annotation update)	CC
DE	Myeloid/lymphoid or mixed-lineage leukemia protein 3 homolog (Histone-	CC
DE	lysine N-methyltransferase, H3 lysine-4 specific MLL3) (EC 2.1.1.43).	CC
GN	MLL3	CC
OS	Mus musculus (Mouse)	CC
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	CC
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	CC
OX	NCBI_TaxID=10090;	CC
RN	(1)	CC
RN	CONCEPTUAL TRANSLATION OF 814-956 AND 1377-4213.	CC
RP	Blatter M.-C.;	CC
RL	Unpublished observations (SEP-2003).	CC
RP	(2)	CC
RP	SEQUENCE OF 957-1376 AND 4214-4894 FROM N.A.	CC
RC	TISSUE=Myeloma;	CC
RC	MEDLINE=21574953; PubMed=11718452;	CC
RA	Tan Y.C., Chow V.T.;	CC
RA	"Novel human HALL (MLL3) gene encodes a protein homologous to ALR and	CC
RT	to ALL-1 involved in leukemia, and maps to chromosome 7q36 associated	CC
RT	with leukemia and developmental defects.;"	CC
RL	Cancer Detect. Prev. 25:454-469(2001).	CC
RN	(3)	CC
RP	SEQUENCE OF 1-813 AND 4803-4903 FROM N.A.	CC
RC	STRAIN=C57BL/6J; TISSUE=Embryo;	CC
RC	MEDLINE=22354683; PubMed=12466851;	CC
RA	Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,	CC
RA	Nikaido I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,	CC
RA	Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojohori T.,	CC
RA	Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,	CC
RA	Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,	CC
RA	Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,	CC
RA	Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,	CC
RA	Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,	CC
RA	Granmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,	CC
RA	Kanai A., Kawaji H., Kawasaki Y., Lee Y., Lenhard B., Lyons P.A.,	CC
RA	Kagott D.R., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,	CC
RA	Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,	CC
RA	Nagashima T., Numata K., Okido T., Pavan W.J., Perteau G., Pesole G.,	CC
RA	Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,	CC
RA	Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,	CC
RA	Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,	CC
RA	Sultana R., Takeuchi Y., Taylor M.S., Teasdale R.D., Tomita M.,	CC
RA	Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,	CC
RA	Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang N.,	CC
RA	Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,	CC
RA	Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,	CC
RA	Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,	CC
RA	Hara A., Hashizume W., Imotani K., Iehi Y., Itoh M., Kagawa I.,	CC
RA	Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,	CC
RA	Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,	CC
RA	Birney E., Hayashizaki Y.;	CC
RT	Analysis of the mouse transcriptome based on functional annotation of	CC
RT	60,770 full-length cDNAs.;"	CC
RL	Nature 420:563-573(2002).	CC
CC	-1- FUNCTION: Belongs to the ASC-2/NCOA6 complex (ASCOM), a	CC
CC	coactivator complex of nuclear receptors, involved in	CC
CC	transcriptional coactivation. MLL3 may be a catalytic subunit of	CC
CC	this complex, which weakly methylates Lys-4 of histone H3. This is	CC
CC	a specific tag for epigenetic transcriptional activation (By	CC
CC	similarity).	CC
CC	-1- CATALYTIC ACTIVITY: S-adenosyl-L-methionine + histone L-lysine =	CC
CC	S-adenosyl-L-homocysteine + histone N(6)-methyl-L-lysine.	CC
CC	-1- SUBUNIT: Belongs to the ASC-2/NCOA6 complex (ASCOM), which	CC
CC	contains ASC-2/NCOA6, the retinoblastoma-binding protein RBQ-3/	CC
CC	RBPF5, alpha- and beta-tubulins, the trithorax group proteins MLL2	CC
CC	and MLL3, and ASH2/ASCL2. Interacts with histone H3 (By	CC
CC	similarity).	CC
CC	-1- SUBCELLULAR LOCATION: Nuclear (probable).	CC

-1- DOMAIN: The SET domain interacts with histone H3 but not H2A, H2B and H4, and may have a H3 lysine specific methylation activity (By similarity).

-1- SIMILARITY: Belongs to the transcription factor trithorax family.

-1- SIMILARITY: Contains 1 DHHC-type zinc finger.

-1- SIMILARITY: Contains 6 PHD-type zinc fingers.

-1- SIMILARITY: Contains 1 post-SET domain.

-1- SIMILARITY: Contains 1 RING-type zinc finger.

-1- SIMILARITY: Contains 1 SET domain.

-1- CAUTION: The regions from 814 to 956 and 1377 to 4213 were deduced from the genomic sequence by similarity to the human sequence.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AY036886; AAK70213.1; -	EMBL; AY036887; AAK70214.1; -
EMBL; AK044828; BAC32109.1; -	EMBL; AK054270; BAC35712.1; -
EMBL; AK077567; BAC36867.1; -	MGI; MGI:2444959; M113.
InterPro; IPR000637; AT hook.	InterPro; IPR001965; ZnF_PHD.
InterPro; IPR001841; ZnF_PHD.	Pfam; PF021178; AT hook; 1.
Pfam; PF00628; PHD; 2.	SMART; SM00249; PHD; 3.
PROSITE; PS00354; HMGI_Y; 1.	PROSITE; PS00868; POST_SET; 1.
PROSITE; PS02080; SET; 1.	PROSITE; PS02216; ZF_DHHC; 1.
PROSITE; PS01359; ZF_PHD_1; 5.	PROSITE; PS00016; ZF_PHD_2; 6.
PROSITE; PS00089; ZF_RING_2; 1.	Transferase; Methyltransferase; Chromatin regulator; Activator;
DNA-binding; Nuclear protein; Transcription regulation; Coiled coil;	Zinc-finger; Repeat.
ZN_FING 340	PHD-TYPE 1.
ZN_FING 343	RING-TYPE.
ZN_FING 387	PHD-TYPE 2.
ZN_FING 435	DHHC-TYPE.
ZN_FING 463	PHD-TYPE 3.
ZN_FING 950	PHD-TYPE 4.
ZN_FING 1000	PHD-TYPE 5.
ZN_FING 1077	PHD-TYPE 6.
DNM_BIND 34	A.T HOOK (BY SIMILARITY).
DOMAIN 4762	SET.
DOMAIN 4887	POST-SET.
DOMAIN 1330	COILED COIL (POTENTIAL).
DOMAIN 1743	COILED COIL (POTENTIAL).
DOMAIN 3047	COILED COIL (POTENTIAL).
DOMAIN 3166	COILED COIL (POTENTIAL).
DOMAIN 3224	COILED COIL (POTENTIAL).
DOMAIN 3387	COILED COIL (POTENTIAL).
DOMAIN 970	CYS-RICH.
DOMAIN 384	CYS-RICH.
DOMAIN 1519	PRO-RICH.
DOMAIN 1708	GLN-RICH.
DOMAIN 1831	PRO-RICH.
DOMAIN 2622	ASP-RICH.
DOMAIN 2682	GLN-RICH.
DOMAIN 3022	SET.
CONFLICT 433	NCRC18C -> VSDFL1CF (IN REF. 3; BAC32109).
CONFLICT 675	S -> C (IN REF. 3; BAC35712).
CONFLICT 4713	MISSING (IN REF. 2; AAK70214).
SEQUENCE 4903 AA; 540182 MW; 0B896490B081BA6C CRC64;	

Db	3291	TSLTMSQP--NFMVMPQLOHQCHTAVISGHTSPARMPSLPGWQSNAS-----	3337	OS	Homo sapiens (Human).
Qy	1731	IIDLSQVPHLPVLVP-----PTGCTATAMDRLAYLPTAQPPSSRHSSS 1775		OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Db	3338	-----AHLPLNPRPTOPPIAQLSLKCTCTPACTVSSANPQ-----NGPPPRVFDDNN 3385		OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Qy	1776	PLSPGGPHLTLPKTTTSSRER-----DRDRDRDREREK-----SILSTTTTVEH 1823		OX	NCBI_TaxID=9606;
Db	3386	PFSESQERERKERLREQEQRVOLMQEVDQRALQORMEQHCLMGAEALANRPVSQ 3445		RN	[1] SEQUENCE FROM N.A. (ISOFORM A).
Qy	1824	APIWRPTEQSGSGSGSGGGGSSRPA-----SHSHAHQHSPI----- 1863		RP	SEQUENCE FROM N.A. (ISOFORM A).
Db	3446	MPFY-----GSDRPFCDLPQPRFLQSPQHQIIGVLOQOQVVOG 3486		RC	TISSUE=Lung;
Qy	1864	---SPRTDALQ---OR---PSVLHN---TGKGIITAVEPSKPTVLRSTSTSGFVRP 1909		RX	MEDLINE=96070776; PubMed=7592926;
Db	3487	SVNSPPNQTQCTNEQRQVGFPSFVDPSPASGSPFNHVKPGHGNLPGSSSQSGLRP 3546		RA	MEDLINE=92049333; PubMed=1719379;
Qy	1910	-----AATFPPA-THCEL-----GGTLGQVYPTLME--PVLLPKAPRVARPERPR 1952		RA	Morinaga T., Yasuda H., Higashio K., Tamaoki T.;
Db	3547	PPTPILPGTSPVANSNVPCGQDPAVTQGNYSQSSQLIQLYSIIPEEKGGKTRTKKK 3606		RA	Morinaga T., Tam T., Ido A., Morinaga T., Miki T., Hashimoto T.,
Qy	1953	ADTHAFLAKPPARGSLPASPSPKSGSEPRPLVPVSGHATIAITPAKNLAPHASDP- 2011		RA	Tamaoki T.;
Db	3607	-----DDAESGKAPSTPHSDCA--APLTGCLSETTTPAVSPSELPOQRQEPV 3656		RT	"Cloning and characterization of an ATBF1 isoform that expresses in a
Qy	2012	-----PAPPASADPHREKTSKPPS--IOELELRSLGYHSGSYSPGCVPSVSPSL 2064		RT	neuronal differentiation-dependent manner.";
Db	3657	EPVPVPTNVASGQCIENSEKLPNSEFIKETSNOQTHVNAEADKPS-----VETPNK 3709		RL	J. Biol. Chem. 270:26840-26848 (1995).
Qy	2065	THDKGLPKHLELDKSHLEGEURPKQGVK--LGGEAAHLPLRLPLPESOPSSPL-LQ 2121		RN	[2] SEQUENCE FROM N.A. (ISOFORM B).
Db	3710	TEBEIKLEK--AETQPSQEDTKVEERTGNKIKDIVAG-----PVSSIQCPHPVGT 3758		RP	TISSUE=Hepatoma;
Qy	2122	TAPGVKHQVTVLAQHI-----SEVITQDYTRHHQPLSAPLAPLYSPFGASCPVLD 2175		RC	MEDLINE=99425270; PubMed=10493829;
Db	3759	TKGDTGNE-----LLKHLKNKASLLTQ----- 3784		RX	Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
Qy	2176	LRPPSDLYLPPDHGAPARGSPHSEGGKRSPEPKNTSVLGGEGD--IEPVSPGEMTE 2233		RA	Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
Db	3785	-----KPEGLTSDSESTKDGKLEIKQSPAEGLQT 3814		RA	Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
Qy	2234	PHSRNAVY-----PLLYRDGEOTEPSPRMGSKSPGNTSOPPAFFSKLTESNAMYKSKQ 2288		RA	Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
Db	3815	LGAQMCGGGGNSQLPKTDG--ASENKKORSKRTQRTGEKAAPRSKRRKD---EBEQ 3869		RT	"Genome duplications and other features in 12 Mb of DNA sequence from
Qy	2289	EINKKLNTHNRNPEYNIQGTETIFNMPAITGTGLMTRYSOAQVHASTNMGLEAI--- 2345		RL	human chromosome 16p and 16q.";
Db	3870	AMYSDDSFTHLQONNLNPNPTPPASLPP-----TPPPMACQRMANGFATTEELAGK 3922		RN	Genomics 60:295-308 (1999).
Qy	2346	-----IRKALMGKYDQWESP--PLSANAFNPLNASASLPAAMPITAADGRSDHTLT 2395		RP	SEQUENCE OF 1151-3703 FROM N.A. (ISOFORM A).
Db	3923	AGVLVSHEVARALGPKPQLPFPFQDDLLARAIAQGPKTVDVPSLPTPHNNHELRQ 3982		RX	MEDLINE=99425270; PubMed=10493829;
Qy	2396	SPGGGKAKVGRPSSRSKAPAGLASGDRPPSVSVHSEGDGNCNRTPLTRVWEDRPS 2455		RA	Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
Db	3983	DHYGDRTPDSFVPS-----SPESVGVVEV-----NKYPDLSLVKEEPE 4023		RA	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
Qy	2456	SAGSTPPFYNPLI-----MRLQAGVMASPPPPGLPAGSGPLAG 2493		RA	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
Db	4024	PVPSPIIPILPSIGKNSRRNDIKTEPGTLFFTSFPG--SSPGRSG 4071		CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RESULT 31				CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
ABF1 HUMAN				CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
ID	ABF1	HUMAN	STANDARD;	CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
AC	Q15911		Q13719;	CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DT	16-OCT-2001		(Rel. 40, Created)	CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DT	16-OCT-2001		(Rel. 40, Last sequence update)	CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DT	10-OCT-2003		(Rel. 42, Last annotation update)	CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
DE	Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)			CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
GN	ATBF1			CC	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.

QY 956 SPQKPLDL-KQLKORAAAIPPIQVTKVHEP-----PREDAAPTKP-----APPAPP-----PP 1002
Db 2386 MPSQAYSAPASANTASSAFLQLTAEBELATFNSKTEAGDEKPKLAEPASQPNQTOE 2445
QY 1003 QNLOPESADPOQGSFGRKSRSPAPPADKFAFAEAQKLPDGPWCWTSGLPVPVPREV 1062
Db 2446 KQGPKEPELQOO--EQPEQKNTNP-----QOKLP----- 2472
QY 1063 IKASPHADPSAFSYPGPHPLGLHDTARPVLPRPPTISNPP--PLISSAKHPSVLER 1120
Db 2473 -----QVLSPLSPQPPQAPPPQCPQLPQSSPSPQLS- 2505
QY 1121 QIGAISQMSVOLHPVYSE-HAKAPVGVPTWG-----LPLPMDPKLA--PFSGVKQEQ--- 1171
Db 2506 -----HLPLKPLHTSTPQOLANLPQOLIYPQCDQCKLAFPSFEHMQEHOOLH 2552
QY 1172 -LSRPGQAGPPE-----SLGVP-----TAGEASVLRGALTALGSPGSGITKGPSTRVPS 1219
Db 2553 FLQAQNPFIHQFLDRSLDMFMLFDPSNPILLASQLLSGAIPQIPASSATS--PST--PT 2608
QY 1220 DSAITYRGSITHGTPADVLVYKGTITRIIGEDSPSRLDRGREDSLPKG-HVIYEGKKGHVL 1278
Db 2609 STMTNLKRLKLEKASAS--PGENDSGTGGEPPQDKRLTITITPEOLEILYQ---KYL 2662
QY 1279 SYEGGMSVTQCKEDGRSSGPPHETAAPKKTYDM-----MEGRVGRASSASIEGLMG 1332
Db 2663 D-----SNPTRKMLDIAHEVGLKKEVQVWFQNTRARERKQGF----- 2701
QY 1333 RAIPP-----ERHSPH-----HLKEQH-----HIR 1352
Db 2702 RAVGPAQAHRCCPFCALFKAKTALAHIRGRHWEAKRAGNYLTLSNMLDCCGLQMK 2761
QY 1353 GSITGIGIPRSYVEAQEDYLRREAKLLKREGTPPPPPSRD-----LTEAYKTQALGPLKL 1407
Db 2762 GDIFDGTFSHL-----PPSSSDGQGVPLSPVSKTMBLSRPTL 2799
QY 1408 KPAHEGLVATVKEAGRSIHEIPRELRHTPPELPLAPRLKESITQGTPLKYDTCASTTG 1467
Db 2800 LSP-----SSIKVEGIEDFESPS-----MSSVNLNFDQTKLDNDCCSVNTAITDTTGDG 2851
QY 1468 SKKHVRSILGSPGRTFFPVHPLDVMDARALERACYBESLKRSGRTASSGSGSIARGAP 1527
Db 2852 NADND--SATGTATETKSSANEGULTKAMMANSEYEDRLSS--GLVSPA----- 2898
QY 1528 VIVPELGPQSPGLTYBHGAFAGHLPRGSPVTVREPTPRLOEGSLSSSKASQDR----- 1583
Db 2899 -----PSFYSKEYDNEGT--VDYSETSSLADPCSPSPG--ASGSACKSGDSDGRPGQK 2947
QY 1584 ----KLSTPREIAKSPHSTVPEHHPHPIPSVYELLRGVGVDLVYRSHIPLAFDPTSIPR 1639
Db 2948 RFRQTMTNLOLVKSCFND-----YRPTMLECEVLGNDI 2983
QY 1640 GIPLOAAAAAY-----LPRHLAPNPT-YPHLYPYLIRGYDPDTAALENRQTI 1685
Db 2984 GLPKRVQVWFQONARAKEKSKLSMAKFGINGQTSYEGPKTECTLCGIKYSARLSVRDHI 3043
QY 1686 -----INDVITSQ-----QMHNTATA---MAQR-ADMLRGLSPRESSLALNYAAGR 1729
Db 3044 FSQHIKSKVDITGSLDKKEYFDPATVRLMAOQELDRI-----KXANEVLGLAAQOO 3098
QY 1730 GIIDLK--QVPHLPVLVPTPGTPATAMDRLAYLPTAPQPFSSRHSSPLSPG---GPTH 1784
Db 3099 GMDNTPLQALNLFTAYPALOGIIPVLLPGL-----NSPSLGFPTPSNTA 3143
QY 1785 LTKPTTTSSSERDRDRDREREKSIILSTTTVEHAPWRPCTEQQSSGSSSGGG 1844
Db 3144 LTPSKPNLNG-----LPSTT-----VPSPLG-PTSLGNPKPSSA 3176
QY 1845 GSSSRPASHAHQHSPISPRTQDALQORPSPVLHNTGMKGIITAVEPSKPTVLRSTSTS 1904
Db 3177 SLSSSPTPAQATNA--MGPOQPPQOOQOOQOQPV-----QQPPPP----- 3213

QY 1905 SPVRPAATPPATHCHGLGTLGDVYPTLMEPVLVLLKPEARVARPRPRADTGHAFIAKPP 1964
Db 3214 ----PAAQPPPTPQLPL-----QQOQKDKDSEKVEKEKAHKGKGP-LPVPK 3258
QY 1965 ARSGLPEASPSKSGSEPRPLVPVPGHATIAATPAKNLAPHHA-SFDDPPAPPASADPHR 2023
Db 3259 KEKGEAPTATAATISAPLPTM-----EYAVDPAQLQALQAAALTSDDPTALLTSQFLPYF 3311
QY 2024 EKTQKPFISIQELRLSLGYHSGSSYSPGVEVSPVSSPSLTHDKGLPKHLELDKSHLE 2083
Db 3312 VFGFSYYPAPQIPGALQSGYQLOFMYGMEGLFFYSPALSOAL---MGL----- 3355
QY 2084 GELRPQKPGVNLKGLGGAHLPHLRPLPESQPSSSPLLQTPAGVKGHQRVVTLAQHISEVI 2143
Db 3356 -----SPGSL-----LQYQYQYQSLQEAIQ-----QQOQRLQOQOQ--- 3388
QY 2144 TDYTRHHPQOISAPLAPLAPLYSPGASCPVLDLRRPPSLLYLPDPDHGAPARGSPHSEGG 2203
Db 3389 -QKVQOQPKASQTPVP-----PGA-----PSPDKD-PAKESPKPEQ 3424
QY 2204 KRSPE-----PNKTSVLGGEDGI-EPVSPPE-----GWTPEGHSRSV 2241
Db 3425 KNTPREVSPLLPKLPEPEABESKASDSLDPFIVPKVQYKLVCKKQAGFSDEEAARSHL 3484
QY 2242 YPLLVRDGEQTEPSRM-----GSKSPGNTSQPPAPFASKLITESNAVVK 2286
Db 3485 KSLCFFGQSVVNLQEMVLHVPTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 3544
QY 2287 KQINKKLNTHN-----RNEPYNISQPGTEIFNMPAITGTGLMTYRS----- 2329
Db 3545 SOHLESALHKHRTITRAARNAKEHPSLLPHSACFPDPFSTASTSQSAAHNSDPPPPSAAA 3604
QY 2330 -QAVQEHASTNMGLEATIRKALMGYDQWEEPSPLSANAFNPLNASASLPA----- 2379
Db 3605 PSSASPHASKSPQVVSRA-----AAKPP-----SPPLSSSTVTSSSCSTSGVQ 3652
QY 2380 -AMPITAADGRSDHTLTSPGGGKAKVSG 2407
Db 3653 PSMTDDYSESDTDLQSKSDGPASPVEG 3681

RESULT 32
TCOF HUMAN
ID TCOF HUMAN STANDARD; PRT; 1411 AA.
AC Q13428; Q99408; Q99860;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Treacle protein (Treacher Collins syndrome protein).
GN TCOF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96154183; PubMed=8563749;
RA Dixon J., Edwards S.J., Gladwin A.J., Dixon M.J., Loftus S.K.,
RA Bonner C.A., Koprivnikar K., Wasmuth J.J.;
RA "Positional cloning of a gene involved in the pathogenesis of
RT Treacher Collins syndrome";
RL Nat. Genet. 12:130-136 (1996).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=97228900; PubMed=9074926;
RA Dixon J., Edwards S.J., Anderson I., Brass A., Scambler P.J.,
RA Dixon M.J.;
RT "Identification of the complete coding sequence and genomic
RT organization of the Treacher Collins syndrome gene";
RL Genome Res. 7:223-234 (1997).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=97250498; PubMed=9096354;

1065 ASPHA-----PDPSAFSYAPPGLHPLGLHDTARVLPVPRPTISNPPPLISSAKHPSVL 1118
 455 ATPSAQVQKWEEDSESS-----EESDSDSGEVPTAVAP-----AQ 491
 1119 BROICATISQGNVOLHVPYSEHAKAP--VGPVTMGLPL--PMDPKLAPFSGVKQEQLS 1173
 492 EKSIGNLOAK-----PTSPAPKPPQKAGPVAQVKAEPKMDNSE-----S 533
 1174 PRGQAGPESLGVPTAOEASVIRGALSGVPGSGITKGIPTSTRVPDSATYRGSITHGT 1233
 534 SEESDSADSEEAAMTAQAQ-----PALKIPOTKACPKKNTTASA 577
 1234 PADVLYKTTIRIIGESPSRLDGRDLSLPGHVIYEGKGVLSYEGGMSV-----TQC 1289
 578 KVAPVRVGT-----QPPR-----KAGTATSPAGSSPAVAGTQR 611
 1290 SKEDGSSSGPPHETAAPKRTYDMWGR-----VGRAI-----SSASIEGLMGRAIPIPERHSP 1342
 612 PAED-----SSSEESDEEKTGLAVTVGQAKSVGLQVKAASVPVKGSLGGTAPV-----665
 1343 HHLKEQHIRSGITOGIPRSYVEAQEDYLRREAKLLKREGTPPPPPSRDLTEAYKTOAL 1402
 666 -----LPKGTGPTVQVKAQED-----SESSEESDSDEAAAPQVKTYSVKTKQA- 713
 1403 GPLKLPKPAHEGLVATVKEAGRSIHEI PREELRHTEPLAPRPLKEGSIQTGTLKYDTG 1462
 714 ---KANPA-----AARAPAKGTIS--APGKVVA 738
 1463 ASSTGSKKHVRSILIGSPRTPPVH-----PLDVMADARALERACYEESLKSR 1511
 739 AAQAKR-----SPSKVPPVNPQNSTVLARGPASVPSGKAVATAA-----QAQ 784
 1512 POTASSSGSIARGAPVIVE--LGKPROSLITYEDHG--APFAGHLPGSPVWRETPR 1568
 785 TGPEEDSGSSEESDESEAEETLAQKPSGKTHQIRAAALP-----AKESPRKGAAPTTP 839
 1569 LEQSLSSSKASQDRKLSTPREIAKSPHSTVPEHHPHIPGYEHLHGVSGVDLYRSHI 1628
 840 GKTGPSAAQKQDSS--GSSSEESDSDEAPAAVTSQVIRP-----880
 1629 PLAF-DPTSIPIRG-----IPLDAAAAYLPHRLAPNPTYPHLYPPVLIYGPDPATALENROT 1684
 881 PLIFVDPNRSAGPAATPAQAQA-----STPRKARASESTARSSSES 924
 1685 INDIYTSQ-----MHNTAT-----AMAQADMRLRGLSPRESSLAIN--YAAGPRGI 1731
 925 EDEDVIPATQCLTPGIRTNVVTMTAHPRIAPKSMAGASSKESRSISDGKKQEGPATQ 984
 1732 IDLSQVPHLPVLVPPPTPGTPATAMDRLAYLPTAQPFFSR--HSSSPILSPGPGTHLTXTPT 1789
 985 VSKKNPASLPL-----TQAAKVLQAQKASEAQPVPVARTQPSGVDSAVGTLPATSPQ 1036
 1790 TTSSSERDRDRDRDRERKSIITSTTVEHAPIWPGTEQSGSGSGSGGGGGSS 1849
 1037 STSVQAKGNTKL-----KPKLPEVQATKAP-----ESSDDSDSDSSSGSEEDGE 1084
 1850 RPASHSHAHQHSPISPRTQDALQORFSLHNTGMKGIITAVFSPKPTVLRSTSTSPVRP 1909
 1085 GPOGAKSAHLTGTPBRTETLVEE-----TAAESSEDDVV-----1119
 1910 AATFPFATHPLGGTLDGVVYPTLMEFVLLPKFAPRVARPERPRADTGHFAFLAKPARSGL 1969
 1120 -----APSQSLLSG-----YMTFGLTP--ANSQASKATPKLD-----SSPSVSSTL 1158
 1970 EPASSPSKGEPRP-----LVPPVSG-HATITARTAKNLAPHASPDPPAPASADPH 2022
 1159 AAKDDPDGQKQEAQPOQAAGMLSPKTKGKEAAGSTTTPKSRKPKKAGNFOASTLALQSMI 1218
 2023 REKTQSPFSGIOELRSLGSLYHGSSYSPGVEPVSPVSPSLTHDKLPKHLFELDKSHL 2082
 1219 TQCLLGQFWPNEAQAQ-----SVMKVLTTEL-----L 1246

2083 EGELRPKQPGPVKLGGEAAHLPHLPLPESQPSPLQITAPGVKHORVVTLAQHISEV 2142
 1247 EQE-RKKVDDTTKSSRKGWESRKRLGQDQAA-----RTPRSKKKKL-----1290
 2143 ITQDTRIHPHPOOLSAPLPAPLYSFFGASCPVLDLRRPFDLYLPPDHGAPARGSPHSEG 2202
 1291 -----GEG 1295
 2203 GKRSPEPNKTSVLGGEDGIEPVSPPEGMTEPHGSRSAVYPLLYRDGQTEFSR-----M 2257
 1296 GEASVSPEKTSITSTGKAKDKAS---GDVKEKKGKS---LGSQGADEPEEELQKGM 1348
 2258 GSKSPGNTSQPPAPFSKLTESNAMYKSKQKQIN--KKLNTNRNEP 2302
 1349 GTVEGDDQSNPKSKKSKKDRKKDKKKEKKKKAKKASTKDSBP 1395
 RESULT 33
 CALL CANFA
 ID CALL CANFA STANDARD; PRT; 1460 AA.
 AC OQXSJ7;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Collagen alpha 1(I) chain precursor.
 GN COL1A1.
 OS Canis familiaris (Dog).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A., AND VARIANT OI ALA-208.
 RC TISSUE=Skin;
 RX MEDLINE=21023337; PubMed=11147834;
 RA Campbell B.G., Wootton J.A.M., MacLeod J.N., Minor R.R.;
 RT "Sequence of normal canine COL1A1 cDNA and identification of a
 RT heterozygous alpha1(I) collagen Gly208Ala mutation in a severe case
 RT of canine osteogenesis imperfecta.";
 RL Arch. Biochem. Biophys. 384:37-46(2000).
 CC -!- FUNCTION: Type I collagen is a member of group I collagen
 CC (fibrillar forming collagen).
 CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
 CC -!- PM: Prolines at the third position of the tripeptide repeating
 CC unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- DISEASE: Defects in COL1A1 are a cause of osteogenesis imperfecta
 CC (OI).
 CC -!- SIMILARITY: Contains 1 VWFC domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; AF153082; AAD34619.1;
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR002181; Fibrinogen_C.
 DR InterPro; IPR001007; VWFC.
 DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; C1g_helix; 2.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWFC_1; 1.
 DR PROSITE; PS0184; VWFC_2; 1.
 DR Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal; Disease mutation.
 FT SIGNAL 1 22 BY SIMILARITY.

Db 943 NSYRLSWGTENDNALSSSIHSGFLVFMWDARGAMRGCRHNL-RIIIPRCKTAP 1001
Qy 240 A-----AHRI-----LEG-----LGPQ-VELPLYNQPSDTRQYH 267
Db 1002 TRVTCRLVKRHLATMPWVEGEGLASRLIEVGPSCAQLFPGVIEIHPF----- 1051
Qy 268 ENIKINQAMRKKLILYFKRNRHARKQWKQKFCORYDQLMLEALEKKVRIENNRRAKES 327
Db 1052 -----AALRGKERELVLRSENGDSWKEHFCDYTEDSELNGLMGDEVLDSPEDLEKK- 1104
Qy 328 KVEYVEKQPEIRKQRELQERMSQSRVQORGSLMSAARSEHEV-----SEIIDGLS 380
Db 1105 RICRIITRDFPOYFAVWSRIKQDNLGPEGVLSSTVVPOVQAFPEGALTKRIRVGLQ 1164
Qy 381 EQ-----ENLEKQMRQLAVIPMLVADQORIKF-----INNGMLMAD-- 418
Db 1165 AQPHESELVKILGNKATFSIV-TLEPRRRKFHKPITMTIPVPKASDVMWNGFGDAP 1223
Qy 419 -----PMKYKO-----RQVMNMWSEQEK 437
Db 1224 TLRLCSITGTTPAQMEDITGTTPLTFVNECVSFTTNVSARFMLIDCRQI-----QES 1277
Qy 438 ETR-----EKFMQHPNFGLIA-- 455
Db 1278 VTFASQVYREIICVPYMAKVFVFAKSHDPIEARLRCFCWTDKVDKTLQOQENFAEVARS 1337
Qy 456 ---SFLERKTV-AEC-----VLYYYLTKNK-----NYKSLVRSY 487
Db 1338 RDVEVLEKPIYVDCFGNLVPLTKSGQHIIFFFAFKENRLPLFVKVRDITQEPGCRUSF 1397
Qy 488 RRRGKSQQQQQQQQQQQQQQMPRSSQBEKDEKEKEKEKEEKEK-PEVENDKEDLL 546
Db 1398 MKEPKS---TRGLVHQALCNLNLITPLIVTSESDQOEEDMTSEKNDTESTETSVL 1454
Qy 547 KE-----KTDDTSGEDNDEKAVASKGRKTANSQGRKGRITRSMANEANSEAIT 597
Db 1455 KSHLVNEVPVLASPDLLSEVEMKQDLIKMTAILTTDVSDBGKAGIKVKELVKAABEEEPGE 1514
Qy 598 PQQSAELASMEI---NE---SSRWTEEMET---AKKGLLEHGRNWSAIAARMVSGKTVS 647
Db 1515 PFEIVERKEDEKNEILRSGTCITRDESSVQSRSEGLVE---BEWVIVS----- 1563
Qy 648 QCKNFYNYKKRONLDELIOQHLKMEKERNARKKKKAPAAASFEAFPPWVE---DEE 704
Db 1564 -----DEBIE-----ARQKAPLEITEY---FCVEVRIDKE 1591
Qy 705 MEASGVSGNEEMVEBAALHASGNEVPRGCSGPATVNNSSD---TESIPSPHTEAAKDT 762
Db 1592 IKG-----KVEKDSGTGLVNYLTDLINTCVLPKQLOQTQV 1626
Qy 763 GONGPKPPA-TLGADGPPPGPTPRRTSRAPISPTPASEATGAPTPPAPPSPPAPPV 821
Db 1627 DKAGKCEALAVGSRSEKGDIIIPDETQ-----STQKHKPSLIGIKKPV 1671
Qy 822 ---VPKEKEBETAAPVVEGEQKPPAAEELAVDTGKAEEPVKS----- 864
Db 1672 RRKLKQKQKEGLQASAEKELKGSSEESLGEDPLAPELPLTVKATSPLEETPIG 1731
Qy 865 -----ECTEAEEGPA-----KGDAEAAEAT-----ABGALKAEKKEGSGR 902
Db 1732 SIKDKVKALQKRVDEQGRSKLPIRVKGDVPKTKTHRPHAPASPSLKSERHAPGSPS 1791
Qy 903 ATTAKSSGAPQDSSATCSADEVDEAGGDKNLLSPRPSLLTPTGDPRANAS---PQK 959
Db 1792 PKTERHSTL-----SSAKTERHPVSPSKTEKHSFVS---PSAKTERHSPASSSKTEKHS 1846
Qy 960 PLDLKQLKQRAAIIPIQVTKV-HEPPREDAAPTKPAAPPAPPPQNLQPSDAPQPCSS 1018
Db 1847 PVSSTKTERHS---PVSSTKTERHPVPSPSGKTDKRPVPSGRT---EKHPPVPSGRT 1900
Qy 1019 PRGKRSRAPPADKEAFAEAQK-----LPGDPPCWTSGLPPPPVPP-----RE 1061

Db 1901 EKRLVPSPSGRTDKHQPVSTAGTEKHLPVSPSGKTEKOP-PVSPTSKTERTIEETMSVRE 1959
Qy 1062 VIKASPHAPDPSAFSAAPPGLPLGLHDHTARPVLPRPPTISNPPPLISSAKHPSVLRQ 1121
Db 1960 LMKAFOSQODPSK-----HKTGLFEHKSQKQPO-----EKQKVRVEKEGKPILTQRE 2008
Qy 1122 IGASIQGMSVOLHVPIYSEHAKAPVGTVMGLPLPMDPKKLAPFSGVKQEQSLSPRQAGPP 1181
Db 2009 -----AQKTENQTIKRGOR----- 2022
Qy 1182 ESLGVPTAQEASVLRGALTALGVPGGSITKGIPTSTPVPSDSALTIRGSIHTGTPADVLKYG 1241
Db 2023 -----LPVTGTAEKRGVRVSSI-----GVK----- 2043
Qy 1242 TITRIIGEDSPSLDRGREDSLPKGHVIEYEGKGHVLSYEGGMSVTCQSKEDGRSSSGPP 1301
Db 2044 -----KEDA-----AGKEKVLSHKIPPEPVQSVPEEESHRESEVP 2078
Qy 1302 HETAPKRTYDMMGRVGRGAISSASIEGLMGRAIPPERHSPHHLKE-----QHHRIGSIT 1356
Db 2079 KERMADEQDMDLQISPRKSTSTDFSEVIKQBELENDKYQQFRLSEETEKAKHLHDQVLT 2138
Qy 1357 Q-----GTPRSYVE-----AQEDYLREAKLLKREGTTPPP-----PPSRDLTEAY 1397
Db 2139 SPFNTPPLDYMKDEFPLALSLQSGALDGSSESLKNEGVAGSPCGSLMEGTPOISSESY 2198
Qy 1398 KTOALGPLKLPKAPAEGLVATVKEAGRSITHEIPRELRHTPELPLAPRPLKEGSIQTGTP 1457
Db 2199 K-----HEGLAETPETSPLSFSPKSEBQTGE-----TKESTKTETTT- 2238
Qy 1458 KYDTGASTGSKKHVRSLSIGPGRTFPVPHPLDVMADARALERACYBESLKSRRPGTASS 1517
Db 2239 -----EIR-----SEKEHPTTKDI 2252
Qy 1518 SGGSIARGAPVIVPELKGPROSPLTYEDHGAPFAGHLPRGSPVNTWRETPRLOE---GSL 1574
Db 2253 TGSSEERGATVEDS-----ETSTESFOKEATLGSP---KDTSPKRQDDCTGSC 2298
Qy 1575 SSSKA-----SODRKLSTPREIAKSPHSTVPEHHPHIPISPYEHLRLGVSGVDLYRSH 1627
Db 2299 SVALAKETPTGLTEAAACDEQRTFGSSAKHTQDSE----- 2335
Qy 1628 IPLAFDPTSPRGIPLDAAAAYILPHRLAPNPTYPHLYPPYLIRGYPD---TAALENRQTI 1685
Db 2336 ---AQESTATS---DETALPLPEASVKTDGTGSKPQGVIRS-PQGLELALPSRDS- 2385
Qy 1686 INDYITSQOMHNTATAMAQRADMLRGLSPRESSALNY-----AAGPGIIDLQOVPHLP 1741
Db 2386 -----EVLASA-----DDSLAVSHKDSLEASP---VLEDNSSHKTP 2419
Qy 1742 VLVPTPTCTPATMDRLAYLPTAPQPSRSHSSPLSP---GG---PTHLTPTTTSSSR 1796
Db 2420 DSLESPSLKESPCRDLSL-----ESSPVEPKMAGIPPSHPLPAAVAKTEL 2465
Qy 1797 ERD-----RDR-ERDRDREREKSILTTTTVTHAPIMWRPGTEQSSGSSGGGGSSSRP 1851
Db 2466 LTEVASVRSRLLRDPDGSAAEDDSLEQTSMB-----SSGKSPLSPDPT 2508
Qy 1852 ASHSHAHQHSPISPRTQDALQOORPSVLHNTG-----MKGIITAVE-- 1891
Db 2509 SSEEVSVE---VTPKTTDVTSPKPAVTHECAEBEDDSENKERRFTPEEMFMVTKIMF 2565
Qy 1892 -----PSKPTVLRSTSTSSPVRPAAT-FPPATHCPGLGTLGDVYVPTLMEPVL 1937
Db 2566 DELEQBAKQKDYKKEPKQBESSSSDDDDDCSVDDVEPKTGTGGEDESGV-----PVL 2619
Qy 1938 LPKEAPRV-----ARPE-----RPRADTGHAFIAPK-----PARSGLEPASSPSKSGE 1980
Db 2620 VTSESRKVSSSESEPELAQLKKGADSG---LLPEPVIRVQPPSPPLPSSMDSNSP---E 2673
Qy 1981 PRPLVPPVSGHATTARTPAKNLAPHHASPPDPPAPASADPHREKTKQSKPSIQBL----- 2036
Db 2674 EVQFPVVSQYTPKMNEDTQEEEPKSEBKSDSHLAEDRHAVSTEAEDRSYDKLNRDT 2733

QY 2037 -ELSLGYHSSSPGCVPS---PVSS-----PSI-----THDKGL 2070
Db 2734 DQPKICDHGC-----EAMSPSSARPVSSGLQSGPTGDDVDEQPVYKESLALQTHEKOT 2789
QY 2071 PKHEELDKHLEGELPKQPGVKLGGEAAHPL-----RPLPSQPSSSPLLQTPAG 2125
Db 2790 EG--EELDVRAES---POADCSESESSSSPLHCLVSGKELDEDSATSIQKTE-- 2842
QY 2126 VKGHRVVTIAQHISEVITQYTHHPQQLSAPLAPLYFPQASCPVLRPPPSDLYL 2185
Db 2843 -----VTKTDTFFNLKDCP---SQDSITTTQTDRFSM---DVPVSDLAENDEIYD 2888
QY 2186 P---PPHGPARGSPHSEGGKSPENKTSVLGGGE-DGIEPVSPPEGMTEPHCHRSV 2241
Db 2889 PQITSPYENVPQSFSSSESKTQTDANHTTSHFSSEVYVTTTSPVEDVVVASSSGTV 2948
QY 2242 YPLLYRDG-----EOTEPS--RMGSKSPGNTSQPPA----- 2270
Db 2949 ---LSKESNFGQDIKMESQLESLWEMQSDSVSSSEFPTMSATTVVGEISKVIITKT 3005
QY 2271 -----PFSSKLTESNAMS---VKSXKQEI-----NKKLTHNRNEPEYNISQPT----- 2311
Db 3006 DVDSDSWSEIRDEDEAFEARVKEBEQKIFGLMVDROSGITPTDTPARTTEEGTPTSEQ 3065
QY 2312 -----EIFNPAITGTLMTYRSQAVQEHASTNMGLEAIRKALMGKYDQWESPPPL 2363
Db 3066 NPFLFQEGKLFEMTR-SGAIDMTKRSYADESFHFQIQES-----REETLSEDKVE 3116
QY 2364 SANAFNPLNASALPA-----AMPITAAD-----GRSDHTLTSPGGGKA 2403
Db 3117 GATGADPLPLETSAESLALSSEKTVDEADLLPDSVSEVEEIPASDAQLSQMG--I 3173
QY 2404 KVSGRPSRKAKS-----PAPGLAGDRPP-----SVSVSESGDCNRRPTLTN 2447
Db 3174 SASSTETPKAVSVGTGDLPTVQTGDIPLPSGVKQISCPDSSEPAVQQLDF---STLTR 3230
QY 2448 RVWEDR---PSSA-----GSTPPFPNPLRMQAVGNVWASPPPPGLPAGS 2488
Db 3231 SVYSDRGDDSPDSPEQKSVIEPTAPMENVPFTESKSKIPVMTPTSTPAPSAEYES 3290
QY 2489 GPLAGPHAWDEPK 2503
Db 3291 SVSEDFLSSVDEENK 3305

RESULT 35
CA17 HUMAN STANDARD; PRT; 2944 AA.
AC Q02358; Q14054; Q16507;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC collagen).
GN COL7A1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=94327588; PubMed=8051117;
RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;
RT "Cloning of human type VII collagen. Complete primary sequence of the alpha 1(VII) chain and identification of intragenic polymorphisms."; J. Biol. Chem. 269:20256-20262(1994).
RN [2]
RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93338437; PubMed=1307247;
RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G., Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E., Uitto J.;

RT "The large non-collagenous domain (NC-1) of type VII collagen is amino-terminal and chimeric. Homology to cartilage matrix protein, the type III domains of fibronectin and the A domains of von Willebrand factor."; Hum. Mol. Genet. 1:475-481(1992).
RL [3]
RN SEQUENCE OF 815-1439 FROM N.A.
RP MEDLINE=91334380; PubMed=1871109;
RA Parente M.G., Chung L.C., Ryyanaenen J., Woodley D.T., Wynn K.W., Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;
RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the gene."; Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
RL [4]
RN SEQUENCE OF 369-1255 FROM N.A.
RP MEDLINE=93107742; PubMed=1469284;
RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisyavanh P.S., Cook M.E., Wright J., Briggaman R.A., Hunt S.W. III;
RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion proteins involved in tissue-specific organization of extracellular matrix."; J. Invest. Dermatol. 99:691-696(1992).
RL [5]
RN SEQUENCE OF 340-675 FROM N.A.
RP TISSUE=Keratinocytes;
RC MEDLINE=92231902; PubMed=1567409;
RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
RT "Molecular cloning and characterization of type VII collagen cDNA."; Biochem. Biophys. Res. Commun. 183:958-963(1992).
RL [6]
RN SEQUENCE OF 2395-2944 FROM N.A.
RP MEDLINE=93271985; PubMed=8499916;
RA Greenspan D.S.;
RT "The carboxyl-terminal half of type VII collagen, including the non-collagenous NC-2 domain and intron/exon organization of the corresponding region of the COL7A1 gene."; Hum. Mol. Genet. 2:273-278(1993).
RL [7]
RN SEQUENCE OF 1-87 FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=94375010; PubMed=8088784;
RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W., Uitto J., Greenspan D.S.;
RT "Structural organization of the human type VII collagen gene (COL7A1), composed of more exons than any previously characterized gene."; Genomics 21:169-179(1994).
RN [8]
RP VARIANT BDDSC ARG-2034.
RX MEDLINE=89227237; PubMed=2653224;
RA Fine J.D., Johnson L., Wright T.;
RT "Epidermolysis bullosa simplex superficialis. A new variant of epidermolysis bullosa characterized by subcorneal skin cleavage mimicking peeling skin syndrome."; Arch. Dermatol. 125:633-638(1989).
RN [9]
RP REVIEW ON DEB VARIANTS.
RX MEDLINE=98041696; PubMed=9375848;
RA Jaervikallio A., Pulkkinen L., Uitto J.;
RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in the type VII collagen gene (COL7A1)."; Hum. Mutat. 10:338-347(1997).
RN [10]
RP VARIANT DEB LYS-2798.
RX MEDLINE=93291877; PubMed=8513326;
RA Christiano A.M., Greenspan D.S., Hoffman G.G., Zhang X., Tamai Y., Lin A.N., Dietz H.C., Hovnanian A., Uitto J.;
RT "A missense mutation in type VII collagen in two affected siblings with recessive dystrophic epidermolysis bullosa."; Nat. Genet. 4:62-66(1993).
RN [11]
RP VARIANT DEB SER-2040.
RX MEDLINE=94224777; PubMed=8170945;
RA Christiano A.M., Ryyanaenen M., Uitto J.;

RT "Dominant dystrophic epidermolysis bullosa: identification of a
RT Gly-->Ser substitution in the triple-helical domain of type VII
RT collagen.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:3549-3553(1994).
RN [12]
RP VARIANT DEB CYS-2623.
RX MEDLINE=96081220; PubMed=8541842;
RA Christiano A.M., Lee J.Y.-Y., Chen W.J., Laforgia S., Uitto J.;
RT "Pretibial epidermolysis bullosa: genetic linkage to COL7A1 and
RT identification of a glycine-to-cysteine substitution in the triple-
RT helical domain of type VII collagen.";
RL Hum. Mol. Genet. 4:1579-1583(1995).
RN [13]
RP VARIANT DEB ARG-2043.
RX MEDLINE=95164985; PubMed=7861014;
RA Christiano A.M., Morriconi A., Paradisi M., Angelo C., Mazzanti C.,
RA Cavallieri R., Uitto J.;
RT "A glycine-to-arginine substitution in the triple-helical domain of
RT type VII collagen in a family with dominant dystrophic epidermolysis
RT bullosa.";
RL J. Invest. Dermatol. 104:438-440(1995).
RN [14]
RP VARIANTS DEB
RX MEDLINE=96220218; PubMed=8644729;
RA Christiano A.M., McGrath J.A., Tan K.C., Uitto J.;
RT "Glycine substitutions in the triple-helical region of type VII
RT collagen result in a spectrum of dystrophic epidermolysis bullosa
RT phenotypes and patterns of inheritance.";
RL Am. J. Hum. Genet. 58:671-681(1996).
RN [15]
RP VARIANT DEB ARG-2575.
RX MEDLINE=96154068; PubMed=8592061;
RA Shimizu H., McGrath J.A., Christiano A.M., Nishikawa T., Uitto J.;
RT "Molecular basis of recessive dystrophic epidermolysis bullosa:
RT genotype/phenotype correlation in a case of moderate clinical
RT severity.";
RL J. Invest. Dermatol. 106:119-124(1996).
RN [16]
RP VARIANT DEB ARG-1782.
RX MEDLINE=96183562; PubMed=8618018;
RA Christiano A.M., McGrath J.A., Uitto J.;
RT "Influence of the second COL7A1 mutation in determining the
RT phenotypic severity of recessive dystrophic epidermolysis bullosa.";
RL J. Invest. Dermatol. 106:766-770(1996).
RN [17]
RP VARIANT DEB ASP-2073.
RX MEDLINE=96310789; PubMed=8757758;
RA Dunnill M.G.S., McGrath J.A., Richards A.J., Christiano A.M.,
RA Uitto J., Pope F.M., Bady R.A.J.;
RT "Clinicopathological correlations of compound heterozygous COL7A1
RT mutations in recessive dystrophic epidermolysis bullosa.";
RL J. Invest. Dermatol. 107:171-177(1996).
RN [18]
RP VARIANTS DEB TRP-1982; GLY-2008; ALA-2025; GLU-2049; TRP-2063 AND
RX ARG-2575.
RX MEDLINE=97465605; PubMed=9326325;
RA Hovnanian A., Rochat A., Bodemer C., Petit E., Rivers C.A., Prost C.,
RA Fraïtag S., Christiano A.M., Uitto J., Lathrop M., Barrandon Y.,
RA de Prost Y.;
RT "Characterization of 18 new mutations in COL7A1 in recessive
RT dystrophic epidermolysis bullosa provides evidence for distinct
RT molecular mechanisms underlying defective anchoring fibril
RT formation.";
RL Am. J. Hum. Genet. 61:599-610(1997).
RN [19]
RP VARIANT DEB ARG-1652.
RX MEDLINE=98106792; PubMed=9444387;
RA Cserhalmi-Friedman P.B., Karpatti S., Horvath A., Christiano A.M.;
RT "Identification of a glycine substitution and a splice site mutation
RT in the type VII collagen gene in a proband with mitis recessive
RT dystrophic epidermolysis bullosa.";
RL Arch. Dermatol. Res. 289:640-645(1997).
RN [20]

RP VARIANTS DEB ARG-2009 AND ARG-2043.
RX MEDLINE=97358588; PubMed=9215684;
RA Winberg J.-O., Hammami-Hausali N., Nilsson O., Anton-Lamprecht I.,
RA Navlor S.L., Kerbacher K., Zimmermann M., Krajci P.,
RA Gedde-Dahl T. Jr., Bruckner-Tuderman L.;
RT "Modulation of disease severity of dystrophic epidermolysis bullosa by
RT a splice site mutation in combination with a missense mutation in the
RT COL7A1 gene.";
RL Hum. Mol. Genet. 6:1125-1135(1997).
RN [21]
RP VARIANTS DEB ASP-1519; ASP-2006; GLU-2015 AND ARG-2034.
RX MEDLINE=98334662; PubMed=9668111;
RA Hammami-Hausali N., Schumann H., Raghunath M., Kilgus O., Luethi U.,
RA Luger T., Bruckner-Tuderman L.;
RT "Some, but not all, glycine substitution mutations in COL7A1 result in
RT intracellular accumulation of collagen VII, loss of anchoring
RT fibrils, and skin blistering.";
RL J. Biol. Chem. 273:19228-19234(1998).
RN [22]
RP VARIANTS DEB CYS-2008; ARG-2207 AND SER-2775.
RX MEDLINE=98410969; PubMed=9740253;
RA Kon A., Pulkkinen L., Ishida-Yamanoto A., Hashimoto I., Uitto J.;
RT "Novel COL7A1 mutations in dystrophic forms of epidermolysis
RT bullosa.";
RL J. Invest. Dermatol. 111:534-537(1998).
RN [23]
RP VARIANT DEB ARG-1347.
RX MEDLINE=99019477; PubMed=9804332;
RA Terracina M., Posteraro P., Schubert M., Sonego G., Atzori F.,
Query Match 3.0%; Score 390; DB 1; Length 2944;
Best Local Similarity 22.0%; Pred. No. 1.3e-05;
Matches 434; Conservative 141; Mismatches 773; Indels 624; Gaps 106;
QY 726 ASGNEVPRGECGSPAT--VNNSSDTESIPSPHTEAAKDTGNGPK-----PPATIGADGP 778
DB 1303 ADGRPGSPGRAGNPGTGCAGLKGSGPLGPRGDPG-ERGRPGKPGCAPQVIGGEP 1361
QY 779 -----PPGPTP-----PR-----RTSRPIETPASEATG 804
DB 1362 GLPRKKGDPGSGPPGPRGPLGDPGPPGPGPLGTAMKDGDRGERGPPGEGGIAPG 1421
QY 805 APTPPPPAPSPSAPPVVPV---KEEKEETAAP-----PVEEGEE---QKPPAAABELAVD 854
DB 1422 EPLGLPLGSGPGQPGVPGPKGKGEKGESEDAFGLPQPGSPGQGRGPPGPGCALGPKGD 1481
QY 855 TGKAEAPVKSECTEAEDEGP-----AKGDAABAAEATAEGALKAEKKEGGGRATTAKSS 909
DB 1482 RG-FPGPL-GEAGEKGERGPPGAGSRGLPGVAGRPGAKG---PEGPGPTGRQCEKGE 1536
QY 910 GAPQSDSSATCSADEVDEAEGDKNRLLSPR-----PSLLTFTGDPANASGP- 957
DB 1537 GRP--GDPVAVVGVAVAGPKGKGDVGP--AGPRGATGVQGERGPPGLVLP-GDPGPKGDPG 1592
QY 958 -OKPLDLKQLKQRAAAATPIQVTKVHEPPREDAAATKPAAPPAPPQNLQPDAPQPG 1016
DB 1593 DRGFIGL-----TGRAGPPG---SGPPGKGDGPRGPPGP-----1626
QY 1017 SSPRGKRSRAPPADKEAFAAEAKLPGDPPCWTGSLPFPVPPREVIVIKASHPADPSAFS 1076
DB 1627 VGPRGRDGEVGEKGDGEG-----PGDP-----GLPGKAGERG-LRGAPGVGRPVGK 1672
QY 1077 -----YAPGHPLPLGLHDTARPVLP-PRPTISNPPPLISSAKHPSVLERQIGAI 1125
DB 1673 GDQGDGPDGDRNGSGSGSGPKG---DRGEGPPGPP-----GRIVDTGPGAR 1716
QY 1126 SQGMSVOLHVYPYSEHAKAPVGPV-TWGLPLPMDPKLAPFSGVKQELSP--RGAGPPE 1182
DB 1717 EKG-----EPGDRGQEGPRGPKDGLPGAPGERGIEGFRGPPGQDQGVGRGAGK 1770
QY 1183 SLGVPTAQEAASVL---RGTAALGVSFPGGSITK-----GIFSTR-----VPSDSAITYRGS 1228
DB 1771 DRGPPGLDGRSLGDKGKGAAGSPGNGAAGKAGDGRDGLPGLRGEQGLPGPSG-----1824

RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., VARIANTS AS GLU-297; ARG-407; ARG-640; ARG-1167;
RP GLU-1207; GLN-1215; SER-1277; THR-1330; GLU-1334; GLU-1347 AND
RP CYS-1661, AND VARIANTS ARG-43; GLU-162; TYR-326; HIS-408; ARG-451;
RP PRO-574; GLU-1269 AND PRO-1474.
RX MEDLINE=21064696; PubMed=11134255;
RA Heidet L., Arrondel C., Forestier L., Cohen-Solal L., Mollet G.,
RA Gutierrez B., Stavrou C., Gubler M.C., Antignac C.;
RT "Structure of the human type IV collagen gene COL4A3 and mutations in
RT autosomal Alport syndrome.";
RL J. Am. Soc. Nephrol. 12:97-106(2001).
RN [4]
RP SEQUENCE OF 1386-1670 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93015826; PubMed=1400291;
RA Quinones S., Bernal D., Garcia-Sogo M., Elena S.F., Saus J.;
RT "Exon/intron structure of the human alpha 3(IV) gene encompassing the
RT Goodpasture antigen (alpha 3(IV)NC1). Identification of a potentially
RT antigenic region at the triple helix/NC1 domain junction.";
RL J. Biol. Chem. 267:19780-19784(1992).
RN [5]
RP SEQUENCE OF 1453-1670 FROM N.A.
RX MEDLINE=91353570; PubMed=1982840;
RA Morrison K.E., Mariyama M., Yang-Peng T.L., Readers S.T.;
RT "Sequence and localization of a partial cDNA encoding the human alpha
RT 3 chain of type IV collagen.";
RL Am. J. Hum. Genet. 49:545-554(1991).
RN [6]
RP SEQUENCE OF 1331-1670 FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=92147878; PubMed=1737849;
RA Turner N., Mason P.J., Brown R., Fox M., Povey S., Rees A.,
RA Pusey C.D.;
RT "Molecular cloning of the human Goodpasture antigen demonstrates it
RT to be the alpha 3 chain of type IV collagen.";
RL J. Clin. Invest. 89:592-601(1992).
RN [7]
RP SEQUENCE OF 1644-1670 FROM N.A.
RC TISSUE=Kidney;
RA Ding J.;
RN Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE OF 1439-1670, AND ALTERNATIVE SPLICING.
RC TISSUE=Kidney;
RX MEDLINE=94124597; PubMed=8294492;
RA Feng L., Xia Y., Wilson C.B.;
RT "Alternative splicing of the NC1 domain of the human alpha 3(IV)
RT collagen gene. Differential expression of mRNA transcripts that
RT predict three protein variants with distinct carboxyl regions.";
RL J. Biol. Chem. 269:2342-2348(1994).
RN [9]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=98196854; PubMed=9537506;
RA Momota R., Sugimoto M., Ohashi T., Kigasawa K., Yoshioka H.,
RA Nimomiya Y.;
RT "Two genes, COL4A3 and COL4A4 coding for the human alphas(IV) and
RT alpha4(IV) collagen chains are arranged head-to-head on chromosome
RT 2q36.";
RL FEBS Lett. 424:11-16(1998).
RN [10]
RP ALTERNATIVE SPLICING.
RX MEDLINE=93280184; PubMed=8505332;
RA Bernal D., Quinones S., Saus J.;
RT "The human mRNA encoding the Goodpasture antigen is alternatively
RT spliced.";
RL J. Biol. Chem. 268:12090-12094(1993).
RN [11]
RP VARIANT PRO-1474.
RX MEDLINE=95078827; PubMed=7987301;
RA Lemlink H.H., Mochizuki T., van den Heuvel L.P.W.J., Schroeder C.H.,
RA Barrientos A., Monnens L.A.H., van Oost B.A., Brunner H.G.,
RA Readers S.T., Smeets H.J.M.;
RT "Mutations in the type IV collagen alpha 3 (COL4A3) gene in autosomal

recessive Alport syndrome.";
Hum. Mol. Genet. 3:1269-1273(1994).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=3;
CC Comment=Additional isoforms seem to exist. Isoforms differ in
CC the C-terminal part of the NC1 domain;
CC Name=1;
CC IsoId=Q01955-1; Sequence=Displayed;
CC Name=2; Synonyms=V;
CC IsoId=Q01955-2; Sequence=VSP_001170;
CC Name=3; Synonyms=L5;
CC IsoId=Q01955-3; Sequence=VSP_001171;
CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
CC colocalized and present only in basement membranes of kidney, eye,
CC cochlea, lung and brain.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the G-
CC X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Isoform 2 contains an additional N-linked glycosylation site.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -!- PTM: Phosphorylated by the Goodpasture antigen-binding protein.
CC -!- DISEASE: Antibodies against the NC1 domain of alpha3(IV) mediate
CC the autoimmune disease Goodpasture syndrome [MIM:233450], which is
CC characterized by hematuria and pulmonary hemorrhage.
CC -!- DISEASE: Defects in COL4A3 are a cause of autosomal recessive
CC Alport syndrome (AS) [MIM:203780], an hereditary disorder
CC characterized by progressive glomerulonephritis, renal failure,
CC hematuria, ocular abnormalities and deafness. The recessive form
CC occurs equally between males and females.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X80031; CAA56335.1; -;
DR EMBL; AJ288487; CAC36101.1; -;
DR EMBL; AJ288488; CAC36101.1; JOINED.
DR EMBL; AJ288489; CAC36101.1; JOINED.
DR EMBL; AJ288490; CAC36101.1; JOINED.
DR EMBL; AJ288491; CAC36101.1; JOINED.
DR EMBL; AJ288492; CAC36101.1; JOINED.
DR EMBL; AJ288493; CAC36101.1; JOINED.
DR EMBL; AJ288494; CAC36101.1; JOINED.
DR EMBL; AJ288495; CAC36101.1; JOINED.
DR EMBL; AJ288496; CAC36101.1; JOINED.
DR EMBL; AJ288497; CAC36101.1; JOINED.
DR EMBL; AJ288498; CAC36101.1; JOINED.
DR EMBL; AJ288499; CAC36101.1; JOINED.
DR EMBL; AJ288500; CAC36101.1; JOINED.
DR EMBL; AJ288501; CAC36101.1; JOINED.
DR EMBL; AJ288502; CAC36101.1; JOINED.
DR EMBL; AJ288503; CAC36101.1; JOINED.
DR EMBL; AJ288504; CAC36101.1; JOINED.
DR EMBL; AJ288505; CAC36101.1; JOINED.

	910	GAPQDSSATCSADEVDEAEGDGKRNLLSPRSILITP-----TGDPANASP-----QKEL	961
QY		: : : :	
Dd	884	PAPPVAISPALTOTP-----LPLPPMAQPQCVLLEDEEFPAPPLTSMQMOL	930
	:	:	:
Qy	962	DLKQLKORAAAIPPIQVTKVHPREDAAATKAAPPAPPPP-----QNLT--QPSDAPOO	1014
	:	:	:
Dd	931	YLQQKLQVKQPPTPLLPSVKVQ-----SQPPPLPFPHHPSVOQQLQQQPPPPPPQ	981
	:	:	:
Qy	1015	PGSPPCKSRSPAAPPADKEAF-----AAEAOKLGDDPCWTSGLPFFVPVPREVIK	1064
	:	:	:
Dd	982	PQPPPPQHOPPRRPVHLQMQFSTHIQQPPPPGQGQPPHHPPGCCQPPPPPAKPQQVIO	1041
	:	:	:
Qy	1065	--ASP--HAPOPSAFSYAPPGCH---PLPLGLHD-----	1090
	:	:	:
Dd	1042	HHSHPRHKKSDPY\$----TGHLEAPSPLMIHSPQMSQFQSLSHQSPPNQPVPKKOEL	1096
	:	:	:
Qy	1091	-TARPVLRP-----PTTSNPPPLISSAKHPSVLERQIGALSOGM	1129
	:	:	:
Dd	1097	RASAVWQPQLVVVVEEKIHSPIIRSBPFSRLRPEPP-----KH-ES-	1140
	:	:	:
Qy	1130	SVOLHVYPSEHAKAPGVFTMGFLPMDKPKLAFFSGV-----XOEOLSBRGOAGPPESLG	1185
	:	:	:
Dd	1141	KAPHLPQRPEMK-----FVDVGREVRIPPONAPPGPADKDOKQ-Q-EKTPIVAPKKDLK	1195
	:	:	:
Qy	1186	VPTAQEASVLRGTLGSVPGGISIKGI PST-----RVPSDSALTYRGISITHGTADVLYK	1240
	:	:	:
Dd	1196	IKN-----MGS-WASLVQKHPTTPSSATAKSSSDSFEPFRFAAREKERKALK	1242
	:	:	:
Qy	1241	GTITRIIGEDSPSLRD--RGRESLPGKHVIEYGKGHVLSYEGGMVSUOCCKEDGRSS	1298
	:	:	:
Dd	1243	AQAESAHEKERLQERMRSRED-----	1272
	:	:	:
Qy	1299	GPHETIAARKTYDMGEICRAISSIASIELCMGRAIPPERSHPHLKQEHHINGSITQG	1358
	:	:	:
Dd	1273	RRAEHARRROEQCCOQCRQEQCCCCCAAAAAATPQAQSSQ-----	1316
	:	:	:
Qy	1359	IPRSVVEAQEDYLR-REAULKIRE	1381
	:	:	:
Dd	1317	-PQSVLDQORELARKEQERRERE	1339
	:	:	:
RESULT 39			
HRX_HUMAN STANDARD; PRT; 3969 AA.			
ID	AC Q03164;	Q13743; Q14845; Q16364; Q9UMA3; 01-OCT-1993 (Rel. 27, Created)	
DT	01-OCT-1993	(Rel. 27, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DT	Zinc finger protein HRX (ALL-1) (Trithorax-like protein).		
GN	MML OR HRX OR ALL1 OR TRX1 OR HTRX. Homo sapiens (Human).		
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
NCBI_TaxID	9606;	[1]	
SEQUENCE FROM N.A.	MEDLINE=93046667; PubMed=1423624; Tkachuk D.C., Kohler S., Cleary M.L.; "Involvement of a homolog of Drosophila trithorax by llq23 chromosomal translocations in acute leukemias."; Cell 71:691-700(1992). [2]		
SEQUENCE FROM N.A.	MEDLINE=96290553; PubMed=8703835; Nilsson I., Loechner K., Siegler G., Greil J., Beck J.D., Fey G.H., Marschalek R.; "Exon/intron structure of the human ALL-1 (MML) gene involved in translocations to chromosomal region 11q23 and acute leukaemias."; Br. J. Haematol. 93:966-972(1996). [3] SEQUENCE OF 1-1909 FROM N.A. MEDLINE=93390935; PubMed=8378076;		

QY 654 FNYKKRONLEILOHKLKMEKERNAREKKKAPAAASEEAAFPVVVEDERMEASGVSGN 713
Db 1189 CWMKRCQNLQ--WMPKAYLOKQAKAVKKKSKTS-----EKDSSKSSVVKN 1236
QY 714 EEEVVEAEALHASNEVPRCEG-----PATVNNSSDTEIPSPH-----TEAKD 761
Db 1237 ---VVDSSOKTTPSAREDPAPKSSSEPPPKPVEEKSEEGNVGAPGESKQATTPASRK 1293
QY 762 TGQNGPKPATLGADPPPGPPT--PPRRTSRAPTEPTPASEATGAPTPPPAPSPS--- 816
Db 1294 SSKVQSQALVI---PPQPTTGPGRK--EVP-KTTPSPKKKQPPPPSGEQSKQK 1345
QY 817 ---APPVVPKKEEETAAAPVVEGE-----EOKPPA----- 847
Db 1346 KVAPRPSIPVOKPKEKEKPPVKNQENAGTINLSTLNGNSKQKIPADGVHRIIRVDF 1405
QY 848 ---ABE-----LAVDTGAE-----EPVKSECTERAEE 872
Db 1406 KEDCEAENVWEMGGILITSVPIITPRVVCFCLASSGHVEFYVCVCCPEPHFKLEE-NE 1464
QY 873 GPAK-----GKDAEAAEATAE-----GALKAEKKE-- 897
Db 1465 RPLEDQLENWCCRRCKFCHVCGROHQATKOLLECNKCRNSYHPECLGPNYPTKTKKKKV 1524
QY 898 ---GSGRATTAKSSGAPQDSSA-----TCSADEVDEAE 930
Db 1525 WICTKVCVKCGSTTPKGWDAQWSHDFSLCHDCAKLFAKGNFCPLCDKCYDDDDYESK 1584
QY 931 ---GGDKNRLSPRPALLTPT-----GDPANASQPKPLDL 963
Db 1585 MWQCGKCDRWVHSCENLSDMEYELNLNLPESVATCVNCTHERPAEWRLALEXELQISL 1644
QY 964 KQ-----LKQRAAAIPIQVTKVHPREDAAPTKAPAPPQNQLQDES 1009
Db 1645 KQVLTALNRSRTTSHLLYRAAKPPDL-----NPETEESIPSRSSPEGPPPPVLTVEYSK 1699
QY 1010 DAPQPGSSPRGSRSPAPPADKAEFAAEAKLPGDPCWTSGLPFPPPREVILKASPHA 1069
Db 1700 QDDQPP-----LDLEGVRKMDQGNYSVLEFSDDDIVIKIQAALNS 1740
QY 1070 PDPFAFSYAPPGHPLPLGLHDTARP-----VLP-----RPPTISNPPPLISS 1111
Db 1741 DG-----GQPEIKKANSMVKSFFTRQMERVFPWFPSVKSRFWENKYVSSNGMLPN 1791
QY 1112 AKHPSVL-----EROIGA-ISQMSVQLHVPYSEHAKAPVGPVTWGLPLPMDPKLA 1162
Db 1792 AVLPPSLDHNVAQWQEREENSHTEQPLMKKIIP-----APKPGGPEPDSPTPLHPPTP 1847
QY 1163 PFGSVKQRLSPRQAGPPESI-----GVPTAQEA----- 1192
Db 1848 ILSTDRSDSP--ELNPPGIEDNRQCALCLTYGDDSDANDAGRLLYIGQNEWTHVNCAL 1905
QY 1193 ---SVLRGTALG---SVFGSITKIGIPSTRVPSDSAITYRGSIT 1230
Db 1906 WSAEFVEDDDGLKNVHMAVIRGQLRCEFCQKPGATV--GCCLTSCTSN----- 1953
QY 1231 HGTPADVLYKGTIYR--IIGEDSPSLDRGREDSLPKGHI-----YEG 1272
Db 1954 ---YHFMCSRAKNCFVLDKVKYQBRHD--LIKGVVPENGFEVFRFVDFEFG 2003
QY 1273 ---KKGHVLSEYEGMSVT-----CQCKEDGRSSSGPP 1301
Db 2004 ISLRKFLNGLEPENIHMIWSMIDCLGILNDLSDCEDKLPFGYQCS----- 2052
QY 1302 HETAAPKRYDMEGRVGRAISSASIEGLMGRAPPPERHSPHLKQEHHRGSIQTQGIPIR 1361
Db 2053 ---RVYWSITDARKCVVTKIVECRPPVVEPDINSTVEHENDRTIAHS-----PT 2100
QY 1362 SYVEAQEDYLREAKLLKREGTPPPPPPSRDLTEAY-----KTQALGPLKLPKPAH 1411
Db 2101 SFTSSSKESQNTABIIIS-PPSPDRPHSQTSGSYHVHISKVPIRPTSPYSPTQSRSGC 2159

QY 1412 EGLVATVKEAGRSIHEIPREELRHTPELPLAPRPLKEGSIQTGTPLYKYDTGASTTSGSKCH 1471
Db 2160 RPL-----PSAG-----SPYPTTHEIIVTGDPL--LSSGLRSIGSRRH 2195
QY 1472 DVRSLIGSPQRTFPVPHPLDVMADARALERACVEE--SLKSRPPTASSGSGSARGAPVIV 1530
Db 2196 STSSL--SPORS-----KLAIMSPMRTGNTYSRNNVSVSTGTAT---DLESSAKVVD 2244
QY 1531 PELCKPQSPPLTYEDHCAPEA-----GHLPRGSPVYMRPPTPRLOQSGLSSS 1577
Db 2245 HVLG-PLNLSSTLQNTSTSNLQRTVVTVGNKNSHLDGSSSEMOKOSSADLVSKSSSL 2303
QY 1578 KASODRKLSTPREIAKSPHSTVPEHHPHIPISVEHLLRGVGVLDYRSHIPLAFDPTSI 1637
Db 2304 KGEKTVLSKSE--CSAHNAVGP-IP-KLAQVH---NTTSRELNVSKIGSAEPSSV 2357
QY 1638 PRGIPLDAAYLPRHLAPNPTVPHLYPYLIRGY-----PD----- 1675
Db 2358 ---SFSKEAL---SPPHLH---LRGQRNDRDQHTDSTQSANSSPDETEV 2399
QY 1676 ---TAALNRQTIINDYI-----TSQWGHN----- 1698
Db 2400 KTLKLSMNRSSIIINHEMSSSRDRKQKCKETFKKHSKSFLEPCQVTTGEGN 2459
QY 1699 ---TATAMAQR-----ADML--RGLS-----PRESS 1719
Db 2460 LKPEFMDVLTPEYMGQPCNNVSSDKIGKLSMPGVKAPPQMVESGAKELQAPKRT 2519
QY 1720 LAL-----NVAAGPRGIIDLSQVPHPLVLPVPT--PGTPATAMDRLAYLPTA--POP-- 1767
Db 2520 VKVTLTPLKMNESQSKNALKESS--PASPLQIBESTPTPEPISASENPGDGFVAQSPNN 2577
QY 1768 ---FSSRHSSPLS---PGPTHLTKPTTSSSRERDRDRERDREREREKIL 1815
Db 2578 TSCODSNNYQNLVQDRNMLMDGP---KPOEDGFKRRYPRRSARAR-----SNMF 2628
QY 1816 TSTTTV-----EHAPIWRPGTEQSSGSSGSGG----- 1843
Db 2629 FGLTPLYGVRSYCEEDIPFYSSSTGKKGKRSABGVGDADDLSTDEDDLYYNNFTTV 2688
QY 1844 -GGSSSRPASHS----- 1855
Db 2689 ISSGGEERLASHNLFREEEQCDLPKISOLDGVDGTSSTVATTTRKSSQIPKRNKEN 2748
QY 1856 ---HAHQHSPISPRTOALQORPSVLHNTG 1882
Db 2749 GTENKLTDRPEDAGEKHVTYKSVGHKNPEMDNCHSVRVKTCQGSLEQLSLESS- 2807
QY 1883 MKGIITAVEPSKPT-----VLSTSTSSP 1906
Db 2808 ---RVHTSTPSDKNLLDTYNTLKLSDSDNNNSDDCGNILPDLMDFVLKNTPSMQA 2862
QY 1907 V--RPAATFPFATHCPLGGTLD---GVY-----PTLMPEV----- 1936
Db 2863 LGESPSSSELLNLGELGLDLSNRKDMGLFEVFSQOLET--TEFPVDSVSSSISABEQF 2921
QY 1937 LLPKEARVARPERPRADTGHAFKAPPARSGLEPASSPSK----- 1977
Db 2922 ELPLIEP-----SDLSVLTTRSTPTVPSONPSRLAVISDSGEKRVITEKSV 2967
QY 1978 ---GSEPRPLVPPVSGHATIARTPAKNLAPH-----ASPPPPAPPASADPHR 2023
Db 2968 ASSESDPALLSPGVD-----PTPEGHMTPDHFIOGHMDADHISPP---PCGSVEQHG 3017
QY 2024 EKTOSKFPSTQELBRLSLGVHSGSYSPGVE--PVSVPVSSPSLTHDKGLPKHLEBDSKHL 2082
Db 3018 NN-----QDLTRNS-----STPGLQVPVSP--TVFIQNKYVPSND----- 3052
QY 2083 EGELRPKQPGVKLGGAHUL--PHLRPLPES---QPSSSPL--LOTAP--GVKGHQVW 2133
Db 3053 ---SPGPSQISNAVAQTTPPHLKPATEKLIWNQNMQPLVQLTFLNGVTKQLT 3105
QY 2134 TLAQHISEVITQDYTRHHQO-----QLSAPLPAFLYSPFGASCPLV----- 2174

Db 3106 SVSSTPVSMTNTSVLPGMGGLTLTGLNPSLTSOSLPSASKGLLPMHSHHQLHSF 3165
QY 2175 -----DLRRPPSLYL---PPDHGAPGSPH-----SEGKRSPP--- 2207
Db 3166 PAATQSSPPPNISNPPSGLLIGVQPPDPOLLVSSSQRTDLSVTVPSSGLKKRPISR 3225
QY 2208 ---EPNKTSLVGGEDGIEPVSPPGMT-----EPCH-----SRSVYPLL 2245
Db 3226 LQTRKNKKLAPGSTPSNIAPSDVSNMTLINFTPSQLPNHPSSLDLGSLNTSSHRTVENI 3285
QY 2246 YRDGEQT-----EPRMGSKSPGNTSQPPAFPSKLTESAMVK-----SKQELNKKLN 2295
Db 3286 IKRSKSSIMYEPAPLFPQSGVGTATAAGTSTISQDTSHLTSQSVSGVGLASSSSVILNVVS 3345
QY 2296 THNRNEPEVNTISQGETEIPFMIPAITGTLMTYRSQAVQEHASTNMGLEAIRIKALMGKYD 2355
Db 3346 MOTTTPPTSSASVPGHVTNPRLLGTDPDGISNLLIKASQOSLGIO-----D 3394
QY 2356 QWEEPPPLSANAFNPLNASALPAAAMPITAADG-----RSDHTLSPGGGK----- 2402
Db 3395 QPVALPP-SSGMFPQLGTSQT-PSTAATAASSICVLPSTQTTGITAASPSGEADHYQL 3452
QY 2403 -----AKVSGRPSRKAKSPAGLSDGDRPPSVSVHSEGDGNCNRTPLNRVWEDRPS 2455
Db 3453 QHVNQLLASKTGIHSSQRLDSASG-----POVSN-----FTQTV--DAPN 3491
QY 2456 SAGSTFPFYNPLIMELQAGVMASSPPPGLPAGSGPLAGPHHAWDEPKP 2504
Db 3492 SMG-----LEON-----KALSSAVQASPTSPG-GSPSPSSGQSRASPSVPGP 3533

RESULT 40
CA54_HUMAN
ID CA54_HUMAN STANDARD; PRT; 1685 AA.
AC P29400; Q16006; Q16126;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 5 (IV) chain precursor.
GN COL4A5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94165049; PubMed=8120014;
RA Zhou J., Leinonen A., Tryggvason K.;
RT "Structure of the human type IV collagen COL4A5 gene.";
RL J. Biol. Chem. 269:6608-6614 (1994).
RN [2]
RP SEQUENCE OF 1-910 FROM N.A., AND VARIANT AS CYS-521.
RC TISSUE=Kidney;
RX MEDLINE=923116923; PubMed=1352287;
RA Zhou J., Hertz J.M., Leinonen A., Tryggvason K.;
RT "Complete amino acid sequence of the human alpha 5 (IV) collagen chain and identification of a single-base mutation in exon 23 converting glycine 521 in the collagenous domain to cysteine in an Alport syndrome patient.";
RL J. Biol. Chem. 267:12475-12481 (1992).
RN [3]
RP SEQUENCE OF 85-1685 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=90337990; PubMed=2380186;
RA Pihlajaniemi T., Pohjola E.R., Myers J.C.;
RT "Complete primary structure of the triple-helical region and the carboxyl-terminal domain of a new type IV collagen chain, alpha 5 (IV).";
RL J. Biol. Chem. 265:13758-13766 (1990).
RN [4]
RP SEQUENCE OF 924-1685 FROM N.A.
RX MEDLINE=91169491; PubMed=2004755;

RA Zhou J., Hostikka S.L., Chow L.T., Tryggvason K.;
RT "Characterization of the 3' half of the human type IV collagen alpha 5 gene that is affected in the Alport syndrome.";
RL Genomics 9:1-9 (1991).
RN [5]
RP SEQUENCE OF 914-1685 FROM N.A.
RX MEDLINE=90160375; PubMed=1689491;
RA Hostikka S.L., Eddy R.L., Byers M.G., Hoeyhtyae M., Shows T.B., Tryggvason K.;
RT "Identification of a distinct type IV collagen alpha chain with restricted kidney distribution and assignment of its gene to the locus of X chromosome-linked Alport syndrome.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:1606-1610 (1990).
RN [6]
RP SEQUENCE OF 1442-1471 FROM N.A.
RX MEDLINE=90252791; PubMed=2339699;
RA Myers J.C., Jones T.A., Pohjola E.R., Kadri A.S., Goddard A.D., Sheer D., Solomon E., Pihlajaniemi T.;
RT "Molecular cloning of alpha 5 (IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus.";
RL Am. J. Hum. Genet. 46:1024-1033 (1990).
RN [7]
RP SEQUENCE OF 1-20 FROM N.A.
RA Guo C., van Damme B., Vanrenterghem Y., Devriendt K., Cassiman J.-J., Marynen P.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBAJ databases.
RN [8]
RP SEQUENCE OF 1258-1270 FROM N.A. (ISOFORM 2).
RX MEDLINE=94133540; PubMed=8301933;
RA Guo C., van Damme B., van Damme-Lombaerts R., van den Berghe H., Cassiman J.-J., Marynen P.;
RT "Differential splicing of COL4A5 mRNA in kidney and white blood cells: a complex mutation in the COL4A5 gene of an Alport patient deletes the NCI domain.";
RL Kidney Int. 44:1316-1321 (1993).
RN [9]
RP REVIEW ON VARIANTS.
RX MEDLINE=97338662; PubMed=9195222;
RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
RT "The clinical spectrum of type IV collagen mutations.";
RL Hum. Mutat. 9:477-499 (1997).
RN [10]
RP VARIANT AS SER-1564.
RX MEDLINE=91169492; PubMed=1672282;
RA Zhou J., Barker D.F., Hostikka S.L., Gregory M.C., Atkin C.L., Tryggvason K.;
RT "Single base mutation in alpha 5 (IV) collagen chain gene converting a conserved cysteine to serine in Alport syndrome.";
RL Genomics 9:10-18 (1991).
RN [11]
RP VARIANT AS ARG-325.
RX MEDLINE=92303559; PubMed=1376965;
RA Knebelmann B., Deschenes G., Gros F., Hors M.-C., Gruenfeld J.-P., Tryggvason K., Gubler M.-C., Antignac C.;
RT "Substitution of arginine for glycine 325 in the collagen alpha 5 (IV) chain associated with X-linked Alport syndrome: characterization of the mutation by direct sequencing of PCR-amplified lymphoblast cDNA fragments.";
RL Am. J. Hum. Genet. 51:135-142 (1992).
RN [12]
RP VARIANT AS GLU-325.
RX MEDLINE=93244772; PubMed=1363780;
RA Renieri A., Seri M., Myers J.C., Pihlajaniemi T., Massella L., Rizzoni G.F., de Marchi M.;
RT "De novo mutation in the COL4A5 gene converting glycine 325 to glutamic acid in Alport syndrome.";
RL Hum. Mol. Genet. 1:127-129 (1992).
RN [13]
RP VARIANTS AS THR-1517; SER-1538 AND GLN-1563.
RX MEDLINE=94010948; PubMed=8406498;
RA Lemmink H.L., Schroeder C.H., Brunner H.G., Nelen M.R., Zhou J., Tryggvason K., Haggama-Schouten W.A.G., Roodvoets A.P., Rascher W.,

RA van Oost B.A., Smeets H.J.M.;
RT "Identification of four novel mutations in the COL4A5 gene of
RT patients with Alport syndrome.";
RL Genomics 17:485-489(1993).
[14]
RN VARIANTS AS GLU-400; VAL-406; VAL-638; ALA-638; ARG-653; ARG-796;
RP ARG-869; ARG-872 AND CYS-1241.
RX MEDLINE=95322976; PubMed=7599631;
RA Boye E., Flinter F., Zhou J., Tryggvason K., Bobrow M., Harris A.;
RT "Detection of 12 novel mutations in the collagenous domain of the
COL4A5 gene in Alport syndrome patients.";
RL Hum. Mutat. 5:197-204(1995).
[15]
RN VARIANT AS ARG-1649.
RX MEDLINE=96213750; PubMed=8651292;
RA Barker D.F., Pruchno C.J., Jiang X., Atkin C.L., Stone E.M.,
RA Denison J.C., Fain P.R., Gregory M.C.;
RT "A mutation causing Alport syndrome with tardive hearing loss is
RT common in the western United States.";
RL Am. J. Hum. Genet. 58:1157-1165(1996).
[16]
RN VARIANTS AS.
RX MEDLINE=96213754; PubMed=8651296;
RA Renieri A., Brutini M., Galli L., Zanelli P., Neri T.M., Rossetti S.,
RA Turco A.E., Heiskari N., Zhou J., Gusmano R., Massella L., Banfi G.,
RA Scolari F., Sessa A., Rizzoni G.F., Tryggvason K., Pignatti P.F.,
RA Savi M., Ballabio A., de Marchi M.;
RT "X-linked Alport syndrome: an SSCP-based mutation survey over all 51
RT exons of the COL4A5 gene.";
RL Am. J. Hum. Genet. 58:1192-1204(1996).
[17]
RN VARIANTS AS, AND VARIANTS ASP-430; SER-444; SER-619; ASN-664 AND
RP MET-1428.
RX MEDLINE=97094179; PubMed=8940267;
RA Knebelmann B., Breillat C., Forestier L., Arrondel C., Jacassier D.,
RA Giatras I., Drouot L., Deschenes G., Gruenfeld J.-P., Broyer M.,
RA Gubler M.-C., Antignac C.;
RT "Spectrum of mutations in the COL4A5 collagen gene in X-linked Alport
RT syndrome.";
RL Am. J. Hum. Genet. 59:1221-1232(1996).
[18]
RN VARIANT AS ASP-1498.
RX MEDLINE=96233932; PubMed=8829632;
RA Tverskaya S., Bobryina V., Tsalykova F., Ignatova M.,
RA Krasnopolkaya X., Evgrafov O.;
RT "Substitution of A1498D in noncollagen domain of $\alpha 5(\text{IV})$ collagen
RT chain associated with adult-onset X-linked Alport syndrome.";
RL Hum. Mutat. 7:149-150(1996).
[19]
RN VARIANT AS GLN-1677.
RX MEDLINE=97295089; PubMed=9150741;
RA Barker D.F., Denison J.C., Atkin C.L., Gregory M.C.;
RT "Common ancestry of three Ashkenazi-American families with Alport
RT syndrome and COL4A5 R1677Q.";
RL Hum. Genet. 99:681-684(1997).
[20]
RN VARIANTS AS ARG-174; ARG-177; ARG-325; CYS-1410; TRP-1421; THR-1517
RP AND ASP-1596.
RX MEDLINE=98112435; PubMed=9452056;
RA Neri T.M., Zanelli P., de Palma G., Savi M., Rossetti S., Turco A.E.,
RA Pignatti G.F., Galli L., Brutini M., Renieri A., Mingarelli R.,
RA Trivelli A., Pinciatoli A.R., Ragaiolo M., Rizzoni G.F., de Marchi M.;
RT "Missense mutations in the COL4A5 gene in patients with X-linked
RT Alport syndrome.";
RL Hum. Mutat. Suppl. 1:S106-S109(1998).
[21]
RN VARIANTS AS VAL-420; 456-PRO--PRO-458 DEL; ASP-573; ASP-624; ASP-635;
RP 802-GLY--PRO-807 DEL; CYS-941; SER-941; SER-1030; SER-1066; ASP-1143;
RP ARG-1196; GLU-1261; SER-1357 AND ARG-1649.
RX MEDLINE=99063529; PubMed=9848783;
RA Martin P., Heiskari N., Zhou J., Leinonen A., Tumelius T., Hertz J.M.,
RA Barker D.F., Gregory M.C., Atkin C.L., Stykardottir U., Neumann H.,
RA Springate J., Shows T.B., Pettersson E., Tryggvason K.;

RT "High mutation detection rate in the COL4A5 collagen gene in suspected
RT Alport syndrome using PCR and direct DNA sequencing.";
RL J. Am. Soc. Nephrol. 9:2291-2301(1998).
[22]
RN VARIANTS AS GLU-579; LYS-633; ASP-947; VAL-953; ARG-1107; ARG-1158;
RP SER-1170 AND TRP-1678, AND VARIANTS SER-444 AND ALA-739.
RX MEDLINE=20030197; PubMed=10561141;
RA Inoue Y., Nishio H., Shirakawa T., Nakanishi K., Nakamura H.,
RA Sumino K., Nishiyama K., Iijima K., Yoshikawa N.;
RT "Detection of mutations in the COL4A5 gene in over 90% of male
RT patients with X-linked Alport's syndrome by RT-PCR and direct
RT sequencing.";
RL Am. J. Kidney Dis. 34:854-862(1999).
[23]
RN VARIANT AS ARG-822.
Query Match 2.9%; Score 379; DB 1; Length 1685;
Best Local Similarity 20.7%; Pred. No. 1.8e-05;
Matches 380; Conservative 136; Mismatches 638; Indels 680; Gaps 90;
QY 764 QNGKPPATILGADGPP-----PGPPTPRRTSRAPIBPTPASEATGATPTPPAPP-- 813
Db 143 QGPPGPGIOMKEGPGOSIMSSLPFGKGNPGYPPGPGIQGLPQGTGIPGIGPPGPG 202
QY 814 -SPSAPPVVV-----PKEEKEETAAPVVEGE--EQKPPAAAE-----L 851
Db 203 MGPPGPGLPKPKGNMGLNFQGPKEGQGLQGPFGPGQISQKRPIDVFEQKQDGL 262
QY 852 AVDTGKAEPVKSECTEAEAGPAKGDAAEAATAEGALKAEKKEGSGRAITAKSSGA 911
Db 263 PGDRGPPGPP-----GIRGPPGPGKEGEGFGKRGKPGKDGNGQPGI---PGL 314
QY 912 PQDSSSATCSADVEDEAGDKN-----RLLSRP----- 942
Db 315 PGDEYGEGERDQ-EGKQGDTPGPPGGLVTPRGTGTIGKGNIGLPLPGEKER 373
QY 943 -----SLLTPTGDP-----RANASPKPLDLK-----QLKORAAAI 973
Db 374 GFPGIQGPPGLPGPPGAAMGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 433
QY 974 P-PIQVTKVHEPPRED-AATKPAAPPAPPPONLOPE-----SDAP 1012
Db 434 PGPEGPAHPHPPSDEICEPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 493
QY 1013 QQPG-----SSPRGKSRSAPPADK-EAFAAAEAOPLGDPDPCWTSGLP-FP---VPPREVI 1063
Db 494 QQPLGPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPPGPP 551
QY 1064 K-----ASHPAP-----DPSAFSY-----APFGHPLPL 1086
Db 552 TFPCKMGDKGELSGFAPGLPGLPTGQDGLPGLPGKPGPPGPPGPPGPPGPPGPP 611
QY 1087 GLHDTARPVLPRPTISNPPPLISSAKHPSVLERQICAIQSGMSVOLHPVSEHAKAPVG 1146
Db 612 GLPGNIGPM--GPPGPPGPPG-----VGEKIGQVA-GNPGQPGIP-----GPKG 653
QY 1147 -----PVTMGLP-----LPMDPKLAPFSGVQEQLSRPGQAGPPESLGVPT 1188
Db 654 DPGQTIQPKPGPLPGNPGRDGDVGLGDDP-----GLPQPGPLPGIPG 696
QY 1189 AQEASVLRTALGSLVPGGSIKGPSTRPVSDSAITVRGSIHTHTPADLVKGTITRIIG 1248
Db 697 SKGEPGPIGLPGPPG--PKGPPGIPGPPGAP-----GTPGRIGLEG----- 737
QY 1249 EDSRSLDRGREDSLPKGHVIECKKHVLSYEGGMSVTCQSKEDGSSSSGPPHETAAPK 1308
Db 738 -----PPGPPGPPGPPG-----EPGFALPGPPGPPGPPG 765
QY 1309 RTYDMWEGRVGRAISSASIEGLMGRAPLPPERHNSPHLKEQHHRGSIQTQIPRSYVAAE 1368
Db 766 -----GFKG-ALGPKGDRGFPPGPP----- 786
QY 1369 DYLRREAKLLKREGTTPPPPPSRDLTBAYKTQALGPLKL-----KPAHEGLVATVKEA 1421

QY 822 VPKEEKEETAAPPVEGEBEQPPAAAEELAVDTGKAEEFVKSECTEEAEEGKAGKDAE 881
Db 46 --KAEKSRQAKKARVESTPK-----ANKGRSEEEISESE--SEETSAKKYKTEE 94
QY 882 AAEATGALKAEKKEGGSRATTAKSGAQQSDSDSATSADVEDEAGGDKNRLLSPR 941
Db 95 LPRQPSDL-----DSLGRSINDGSSDPRDIDQ-----NR--STS 131
QY 942 PSLTPTGDPANASPKQPLDLKOLKORAAAIPIQVTKVHEPPREDAAATKAPAPPP 1001
Db 132 PSIVSP-----GSVENSDSSGLSQGARP-----YHPPPL-----FPPSPPP 170
QY 1002 PONTL--QPESDAPQPGSSPRGKSRSPAPADKAEFAAEQAQKPGDPCWTSGLPF--- 1055
Db 171 PDSIPROFEGFEPHPSVPTFG-YHAPMEPTSLFOGPP--PGAPPPHQLPYGSGAGG 226
QY 1056 -----VPPPREVIKASPHADPSAFSVAPPGHPLPLGLHDTA-----R 1093
Db 227 GVLGPPMGPKGGAASVGPSPGKQHPPTTPIPISSSGASGAPPAKPPNTPVGAGNL 286
QY 1094 PVLPRPTISN-----PPPLISSAKHPSVLEROIGALSQMSVOLHVPYSEHAKA----- 1143
Db 287 PSAPPATFPHVTENLPPPPALRLNNASAPPGMGA-----QPTPGHLP-SPHAMGQMS 341
QY 1144 --PVGPTWGLPLPMDPKLAPFGSVKQEQSLPRGQAGPPESLGVPTAQASVLRGTALG 1201
Db 342 GLPPGP--EKGPTLAPSPHPLPASS-----SAPGPPMRYPYSSCSSSV----- 384
QY 1202 SVPGSITKGIPTSTRVPSDSAITYRGSIHTGTADVLVYKGTITRIIGEDSPSLDRGRED 1261
Db 385 ---AASSSSSAAATQYPASQTLPL--SYPHSFPPL-----TSMVSVNQPKY--TQP 428
QY 1262 SLPKGHVITYEKK-----GHVL-----SYEGMSVTQCKEDGRSSSGPPHETAPARTY 1311
Db 429 SLPSQAVMSQGPPIPPPPYGRLLNNTHPGFFPPT-----GGQSTAHPP-----APAHH 479
QY 1312 DMWEGRVCAISSASIEGLMGRAPPERHSPHLKEQHIRGSITQGTIPRSY---VBAQE 1368
Db 480 HQQQ-----QOPQOPQOQH-HGNSGPPPPGAYPIPLESSN 516
QY 1369 DYLRREAKLKREGTPPPPPSRDLTEAYKTQALGPLKLKPAHEGLVNAVIVKEAGRSIHEI 1428
Db 517 SHHAHPYNNPSLSGLRPYP-----GPAHLPPSH-QGV-SYSQAG----- 555
QY 1429 PREELRHTPELAPRPLKEGSIQGTPLKYDTGASTTGSKKHVRSILIGSGRT--PPP 1486
Db 556 -----PNGP-PVSSSSNSGSSSQAYSCSPSSSQ-----GPGASYPPPP 596
QY 1487 VHLDDVMADARALACVYESLSKSRPGTASSGSGSIARGAPVIVPELGPQSPPLYE-- 1544
Db 597 VPPITTS-----ATLSTVIATVASPGYKKTASPPGPQYSKRAPSPGSKYKTA 645
QY 1545 -----DHGAP-----FAGHLPRGSPVTMREPTPLQEGSISSSKASQDRKLSTTP 1589
Db 646 TPGYKPGSPSFRGTGTPGYRGTSPPAGPGTFKPGSPVTGPGPLPPAGPSSLSLPPPP 705
QY 1590 REIAKPSHSTVPEHHPHISPYEHLRGVSGVDLYRSHIPLAFDPTISIRGIPLDAAAY 1649
Db 706 AAPTGTGPTATQIKQEPAAEYE-----TPESPVPARSPPPKVVD----- 748
QY 1650 YLPHLAPNPTPHLYPPVILRGYPDTAALENQTIINDYITSQQMHNTATAMAQRADM 1709
Db 749 -VPSHASQAR-----FNHLDKGFNSCA-----RSDL 775
QY 1710 LRGLSPRESSLALNAAAGRGIIIDLSQVPHLPVLVPTGTATAMDRLAYLPTAPQPS 1769
Db 776 Y--FVPLEGS-----KLAKKGRADLVEKVRREAE 801
QY 1770 SRHSSSPLSPGPGTHLTKPTTTSSSERDRDRDREREKSIILTSTTIVE--HAPIW 1827
Db 802 QR-----AREEKEREKEREKEREKEREKERSVKLAQEGRAPVE 843

QY 1828 RPTEQSSGSSGSSGGGSSSRPASHAHQHSPISPRTQDALQORQSPVHLNTHMGKII 1887
Db 844 CP-----SLGEPVPH-----RPPPEPGS--AVATVPPYL-GPDTPALR 877
QY 1888 TAVEPSKPTVLRSTSTSSPVRPAATFPATHCHPLG-----GTLGVTPTLMEPVLVLPKEAP 1943
Db 878 TLSEYARPHVMSFGNRNHPF-----YVPLGAVDPGLLGYNPALYSSDPAARE 927
QY 1944 RVARPERPADTCHAFILAKPPARSGLEPASS-PSKGSERPLVPVPSGHATIAARTAKNL 2002
Db 928 REAREDRDLRLKPGFEVXP--SELEPLHGVPGGLDP--FPHGGLALQALQPPGLHP 981
QY 2003 APHASFPD-----PAPPASADPHREKTSKPFISOELRSIGYHSSYSVEGVE 2054
Db 982 FPHPSLPLGERERLALAGPALRPD-----MSYAE-RLAAERQHAERVAALGND 1030
QY 2055 PVSFVSPSLTHDKGLPKHLEELDCKSHLEGELRPQPGPVKLGEAAHLPHLPLPESQP 2114
Db 1031 PLARLOMLNT-----PHHQH--SHIHSHLHQODAIHAASASVH-PLIDPL----- 1076
QY 2115 SSSPLQLTAPGVKHQHVVTLAQ-----HISEVITQDYTRHH-----PQOLSAP 2158
Db 1077 ASSHLTRIPYAG-----TLNPLPLPHLPHENEVL-----RHQLFAAPYRDLPLASLAP 1126
QY 2159 LPAPLYSFPGASCEVLDLRRPPSD-----LYLPPPDHGA--PARGSPHSEGGKRSPEP 2209
Db 1127 MSA-AHQLOAMHAQSAELQRLALEQQOQLHAHHPHLSVPLPAQEDYVYSHLKESDKP 1182

RESULT 42
ABF1 MOUSE
ID ABF1 MOUSE STANDARD; PRT; 3726 AA.
AC O61329;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)
DE (AT-binding transcription factor 1).
GN ATBF1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/MK X ICR; TISSUE=Brain;
RX MEDLINE=96194902; PubMed=8654949;
RA Ido A., Miura Y., Watanabe M., Sakai M., Inoue Y., Miki T.,
RA Hashimoto T., Morinaga T., Nishi S., Tamaoki T.,
RT "Cloning of the cDNA encoding the mouse ATBF1 transcription factor.";
RL Gene 168:227-231(1996).
RN [2]
RP INTERACTION WITH FNBp3
RX MEDLINE=97315177; PubMed=9171351;
RA Bedford M.T., Chan D.C., Ieder P.;
RT "FAP WW domains and the Abl SH3 domain bind to a specific class of
RT proline-rich ligands.";
RL EMBO J. 16:2376-2383(1997).
CC -!- FUNCTION: Transcriptional activator that binds to the AT-rich core
CC sequence of the enhancer element of the AFP gene.
CC -!- SUBUNIT: Interacts with FNBp3.
CC -!- SIMILARITY: Contains 4 homeobox domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch>)
CC or send an email to license@isb-sib.ch.
CC -----
CC EMBL; D26046; BAA05046.1; --


```
Db 2480 -KLQPP-----AAPSILQPPQAPPQCPPLQSSSPSPQLSHLPLKPLHTSTP-----2526
Qy 1037 AEAQKLPDGP-----CWTSGLPFPVPPREVIVKASPHADPPSAFSAVPPG-----HP-----1083
Db 2527 ---QOLANLPPQLIPYQCDCKLAPP-----SFEHQHSHOQLHLFSAQNOQFIHQFQL2575
Qy 1084 ---LPLGLHDTARPLVPRPTISNPPPLI---SSAKHPS-----VLERQI---GAIS1126
Db 2576 DRSLDMFPLDPNPLL-ASQLISGALPQIPASATSPSTPTMTMTLKRKLKLEKASAS2634
Qy 1127 QGMS-----VOLHVPYSEHAKAPGVFTVMGLPLPMDPK-----1160
Db 2635 PGENDSGTGGSEPODKRLRTITITPEOLEILYQYL-----LDSNPRKMLDH2682
Qy 1161 LAPSGVQKEQ-----SPRQO---AGPPES-----LGVPTAQEA-----1192
Db 2683 IAEVGLKRVVQVWFQNTARERKRGQFRAVGAQAHRRCFCRALFAKTALEAHIRSR2742
Qy 1193 -----SVLRGTALGSV-PGGSITKGIP-----1213
Db 2743 HWHEAKRAGYNLTLSAMLLDCDGLQMGKIDIFDGTSPSHLPSSSDGQGVPLSPVSKTME2802
Qy 1214 -----STRVPSDAITYRGSITHGTGPADVLXK1240
Db 2803 LSPRTLSPSSIKVEGIEDFSPSMSSVNLNFDQTKLDNDCCSSVNTAITDTTGGD---E2859
Qy 1241 G-----TITRIIGEDSPRLDRG-----REDSLPKGV-----IVE---GKKG1276
Db 2860 GNADNDSATGIATKSSAPNEGTLKAAMWAMSYEDRLSSGLVSPAPSFYSKEYDNEG2919
Qy 1277 VLSYEGGMSVTCSEKEDGRS-----SSGPHHTAAPK-----RTY1311
Db 2920 VDYSETSLADPCSPSPGASGACKSGDGGRRPQCKRPTQMTNLQLKVLKSCFNDRTP2979
Qy 1312 DMMEGRV-----GRA-----ISSASIEG-----LMGR1333
Db 2980 TMLCEVLNGDIGLPKRVVQVWFQNAKAKESKLSMAKHFGINQTSYEGPKTECTLCGI3039
Qy 1334 AIPPERISPHLKQHHIRGSIOTIGIPRSYVEAOEDYLREAKLLKREGTPPPPPPERDL1393
Db 3040 KYSARLSVRDHIFSQOHI--SKVXDTTIGSLQDKKEYF-----DPAIVRQL3083
Qy 1394 TEAVKTOALGPLKPAHEGLVATVKEAGRSIHBIPREELR---HTPELPLAPRLKEGS1450
Db 3084 MAQELD-----RIKKANEVLGLAAQOOG-MFNAPLOALMLPTTYPALQGIPLVLLPGL3137
Qy 1451 ITQGTPLKYDTGASTGSKKHVRSLLIGSPGRTFP-FVHPLDVMADARALACRYESLX1509
Db 3138 NRPSLPGFTPANTALTSPK---NLMGLPSTTVSPGLP-----TSGLP3178
Qy 1510 SRPGTASSGSGSIARGAPVIVPELGKPROSLTYEDHGAPFAGHLPRGSPVTMREPTPL1569
Db 3179 NKPSASLSSTPAQAATMAMAPQPPQPPQPPVQPPPP-----PPAAQQIPAPQLTPQ3234
Qy 1570 QEGSLSSSKASQDKL---TSTPREIAKSPHSTVPEHHPHPIIS-PYEHLLRGVSGVDLYR1625
Db 3235 QRKDKGEGKEKEKAHKGEKEPLVPKKEGEAPPACTGTISAPLPAVEYVDPQAQLOA3294
Qy 1626 SHIPLAFDPTS-----IPRGIPDLAAAYLYLPHLAPNPTYP-HLYPPYLLIRG-YPDTA1677
Db 3295 LQAALTSDPTALLTSQFLPYFVP--GFSPYVAPQ--IPGALQSGVLPQPMYGMGLFFVSP3350
Qy 1678 AL-----ENRQTINDYITSCQMHNTATAMAQBRADMLRGLSPRESSLA1721
Db 3351 ALSRLMGLSPGSLLOQYQYQBSLQEAIOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQOQ3410
Qy 1722 LNYAAGPRGIIDLSQVPLPLVLPPTCTPATMDRLAYLPTAPOPPSSRH---SSGPLS1778
Db 3411 VQOQOQOQOQKASQTP-----VFQGAASPKDKPAK-----ESPKEQKNVPRELSPLL3460
Qy 1779 PGGPHTLTKPTTTTSSSERD-----RDRERDRERE-----KS1813
```

```
Db 3461 KPPEPEARSKSASADSLCDPFFIVPKVQKLVCRKQACGFGDEAARSHLKSICCFGQS3520
Qy 1814 ILTSTTTTVEHAPITWRPQTEQSSSGSSGGGGSSSRPASHAHQHSPTSPRTQDALQO1873
Db 3521 VVNLOEWLHV---TSGGGGGGGGGGGGGGGGGYHCLACESALCGEALSOHLESALHK3577
Qy 1874 RPSVLHNTGMKGIITAVEPSKPTVLRSTSTSSPVPRPAATFPFATHCPCLGGTLDGVYPTLM1933
Db 3578 -----HRTITRAARNAKE-----3590
Qy 1934 EPVLLPKAEPVARPERPRADT---GHAFAPKPARSGLEGPASSPSKSGSEP---RPLVP1986
Db 3591 HPSLLPHSA---CFPDSTASTSQSAAHNSDPPP-----PSAAPSSASPHASRKSWP3641
Qy 1987 PVSCHATIARTPAKNLAPHHASPDPAPPASADPHREKTQSKPFSIQEELRSLGVHGS2046
Db 3642 PVGRASAAKPPS-----FPPLSSST-----VTSS3667
Qy 2047 SYSPGVEPVSPVSSPSLTHDKGLPKHLELDKSHLEGELRPKQPGPVKLG2098
Db 3668 SCSTSGVQSPMPTDDYSESDTDLQK-SDGPASPVG---PKDPSCFKDSG3715

RESULT 43
CA24 ASCSU
ID CA24 ASCSU STANDARD; PRT; 1763 AA.
AC P2733;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, last sequence update)
DT 15-MAR-2004 (Rel. 43, last annotation update)
DE Collagen alpha 2(IV) chain precursor.
OS Ascaris suum (Pig roundworm) (Ascaris lumbricoidea).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea;
OC Ascarididae; Ascaris.
OX NCBI_TaxID=6253;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS I AND II).
RX MEDLINE=91340768; PubMed=1714907;
RA Pettitt J., Kingston I.B.;
RT "The complete primary structure of a nematode alpha 2(IV) collagen
and the partial structural organization of its gene.";
RL J. Biol. Chem. 266:16149-16156(1991).
CC -!- FUNCTION: Collagen type IV is specific for basement membranes.
CC -!- SUBUNIT: Trimers of two alpha 1(IV) and one alpha 2(IV) chain.
CC Type IV collagen forms a mesh-like network linked through
intermolecular interactions between 7S domains and between NC1
domains.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=I;
CC IsoId=P27393-1; Sequence=Displayed;
CC Name=II;
CC IsoId=P27393-2; Sequence=VSP_001159;
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
domain (NC1) at their C-terminus, frequent interruptions of the
G-X-Y repeats in the long central triple-helical domain (which may
cause flexibility in the triple helix), and a short N-terminal
triple-helical 7S domain.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
are involved in inter- and intramolecular disulfide bonding. 12 of
these, located in the NC1 domain, are conserved in all known type
IV collagens.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC
```


Db 1461 -----GLPGLDGLPPSG-----PPGAGAKGRDGPFGPGMGKEGAP----- 1499

QY 2213 SVLGG--CEDGIEPVSPGEGTEPHGHSAYVPLLYRDG-----EOTESPBMGSKSPGNTS 2266

Db 1500 ----GLPGFPGIEGIPGPGIPGPGSPGPGPGP--SYKDGFLLVKHSOTSEVPQCPGPMVK 1554

QY 2267 QPPAFPSKLTSTNSAMVSKKQ-----EINKLNTNRNEPEYNI 2306

Db 1555 LWDGSLLYIEGNE---KSHQDLGHAGSCLSRSTMPFLFCDDVNNVCNYSARNDKSYWL 1611

QY 2307 SQPGTEIFNMPAITGTGLMYTSQAVQEHASTNMGLEAIIRKALMGKYDQWEESPPLSAN 2366

Db 1612 ST--TAPIPMVPSGGIEPIYSRCAVCEAPAN-----VIAVHSQIQIPN-CPN 1658

QY 2367 AFNPLNASASIPAAPIYTAADGRSDHTLTSPGGGKAKVSRPSRKAKSPAPGLASGDR 2426

Db 1659 GWNLSIWIGYSF--AM-----HTGAGAEQGGQS-----LSSPGSCLEDPA 1696

QY 2427 PPSVSSVHSEGDEN 2440

Db 1697 TPTFECNGARGTCH 1710

RESULT 44

DRPL_HUMAN

ID DRPL_HUMAN STANDARD; PRT; 1185 AA.

AC P54259; Q99495; Q99621; Q9UEK7;

DT 01-OCT-1996 (Rel. 34, Created)

DT 10-OCT-2003 (Rel. 42, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Atrophin-1 (dentatorubral-pallidoluysian atrophy protein).

GN DRPLA.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain; and Cerebellum;

RX MEDLINE=95144175; PubMed=7842016;

RA Nagafuchi S., Yanagisawa H., Ohsaki E., Shirayama T., Tadokoro K., Inoue T., Yamada M.;

RA "Structure and expression of the gene responsible for the triplet repeat disorder, dentatorubral and pallidoluysian atrophy (DRPLA)."; Nat. Genet. 8:177-182(1994).

RL [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Brain;

RX MEDLINE=96026098; PubMed=7485154;

RA Onodera O., Oyake M., Takano H., Ikeuchi T., Igarashi S., Tsuji S.;

RA "Molecular cloning of a full-length cDNA for dentatorubral-pallidoluysian atrophy and regional expressions of the expand alleles in the CNS."; Am. J. Hum. Genet. 57:1050-1060(1995).

RN [3]

RP SEQUENCE FROM N.A.

RC MEDLINE=96262314; PubMed=8965642;

RA Margolis R.L., Li S.-H., Young S.S., Wagster M.V., Steine O.C., Kidwai A.S., Ashworth R.G., Ross C.A.;

RA "DRPLA gene (atrophin-1) sequence and mRNA expression in human brain."; Brain Res. Mol. Brain Res. 36:219-226(1996).

RL [4]

RP SEQUENCE FROM N.A., AND POLYMORPHISM OF POLY-GLN REGION.

RC TISSUE=Brain;

RX MEDLINE=97228904; PubMed=9074930;

RA Ansari-Lari M.A., Shen Y., Muzny D.M., Lee W., Gibbs R.A.;

RA "Large-scale sequencing in human chromosome 12p13: experimental and computational gene structure determination."; Genome Res. 7:268-280(1997).

RL [5]

RN [5]

RP SEQUENCE OF 1-76 FROM N.A.

RX MEDLINE=97005364; PubMed=8852663;

Yanagisawa H., Fujii K., Nagafuchi S., Nakahori Y., Nakagome Y., Akane A., Nakamura M., Sano A., Komure O., Kondo I., Jin D.K., Soerensen S.A., Potter N.T., Young S.R., Nakamura K., Nukina N., Nagao Y., Tadokoro K., Okuyama T., Miyashita T., Inoue T., Kanazawa I., Yamada M.;

RT "A unique origin and multistep process for the generation of expanded DRPLA triplet repeats"; Hum. Mol. Genet. 5:373-379(1996).

RL [6]

RN SEQUENCE OF 470-725 FROM N.A.

RP TISSUE=Brain cortex;

RC MEDLINE=93315145; PubMed=8325628;

RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;

RT "Novel triplet repeat containing genes in human brain: cloning, expression, and length polymorphisms."; Genomics 16:572-579(1993).

RL [7]

RN INTERACTION WITH WWP1 AND WWP2.

RP MEDLINE=98313405; PubMed=9647693;

RA Wood J.D., Yuan J., Margolis R.L., Colomer V., Duan K., Kushi J., Kaminsky Z., Kleiderlein J.J. Jr., Sharp A.H., Ross C.A.;

RT "Atrophin-1, the DRPLA gene product, interacts with two families of WW domain-containing proteins."; Mol. Cell. Neurosci. 11:149-160(1998).

CC -!- SUBUNIT: Interacts with WWP1 and WWP2.

CC -!- TISSUE SPECIFICITY: Relatively high levels in the brain, ovary, testis and prostate. Lower levels in the liver, thymus and leukocytes.

CC -!- POLYMORPHISM: The poly-Gln region of DRPLA is highly polymorphic (7 to 23 repeats) in the normal population and is expanded to about 49-75 repeats in DRPLA patients. Longer expansions result in earlier onset and more severe clinical manifestations of the disease.

CC -!- DISEASE: Defects in DRPLA are the cause of dentatorubral-pallidoluysian atrophy (DRPLA) [MIM:125370], an autosomal dominant neurodegenerative disorder characterized by a loss of neurons in the dentate nucleus, rubrum, globus pallidus and Luys' body.

CC Clinical features are myoclonus epilepsy, dementia, and cerebellar ataxia. Onset of the disease occurs usually in the second decade of life and death in the fourth.

CC -!- CAUTION: Ref.2 sequence differs from that shown due to several frameshifts.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL; D31840; BAA06626.1; -.

DR EMBL; D38529; BAA07534.1; ALT_FRAME.

DR EMBL; U23851; AAB50276.1; -.

DR EMBL; U47924; AAB51321.1; -.

DR EMBL; D63808; BAA23631.1; -.

DR EMBL; L10377; -; NOT_ANNOTATED_CDS.

DR PIR; G01763; G01763.

DR Genew; HGNC:3033; DRPLA.

DR MIM; 607462; -.

DR MIM; 125370; -.

DR GO; GO:0005717; Cytoplasm; TAS.

DR GO; GO:0005634; C.nucleus; TAS.

DR GO; GO:0005515; F.protein binding; IPI.

DR GO; GO:0007417; P.central nervous system development; TAS.

DR InterPro; IPR002951; Atrophin.

DR Pfam; PF03154; Atrophin-1; 2.

DR PRINTS; PR01222; ATROPHIN.

KW Triplet repeat expansion; Polymorphism; Epilepsy.

FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).

FT DOMAIN 302 305 POLY-PRO.

FT DOMAIN 376 382 POLY-SER.

FT DOMAIN 386 397 POLY-SER.

FT DOMAIN 442 447 POLY-PRO.
FT DOMAIN 479 483 POLY-HIS.
FT DOMAIN 484 497 POLY-GLN.
FT DOMAIN 504 507 POLY-PRO.
FT DOMAIN 564 574 POLY-SER.
FT DOMAIN 704 707 POLY-PRO.
FT DOMAIN 802 815 ARG/ALA-RICH (MIXED CHARGE).
FT DOMAIN 816 827 ARG/GLU-RICH (MIXED CHARGE).
FT DOMAIN 925 934 ARG/GLU-RICH (MIXED CHARGE).
FT CONFLICT 94 94 MISSING (IN REF. 3).
FT CONFLICT 333 333 H -> Y (IN REF. 1).
FT CONFLICT 339 339 M -> I (IN REF. 3).
FT CONFLICT 541 541 P -> T (IN REF. 6).
FT CONFLICT 1028 1028 A -> G (IN REF. 1).
SQ SEQUENCE 1185 AA; 124773 MW; 5135744CEE491C18 CRC64;

Query Match 2.8%; Score 368; DB 1; Length 1185;
Best Local Similarity 21.0%; Pred.No. 2.9e-05;
Matches 327; Conservative 139; Mismatches 604; Indels 490; Gaps 74;

QY 762 TQNGPKPATLGADGPPGPPPTPRRTSRAPIEPTPAEATGAPTPPPAPSPSPAPPV 821
DB 3 TRONKDSMSGRKKEAPGREELRSGRASPGCVSTSSDG----- 45

QY 822 VPKEKEEETAAAPPVEG-----EQQPPAAELAVDTGKAEEPVKSECTEABEGPAKKG 878
DB 46 --KAEKSQTAKARVEASTPKVNQKGRSEISESEETNAPKTKTEQELPRQSPS 103

QY 879 DAEAAEATAGALKAEKGGSGRATTAKSGAPQDSSATCSADEV-----DE 928
DB 104 DLDSLD-----GRSLNDGSDPRDIDQDNRTSPSYSPGSEVNDSDS 147

QY 929 AEGGDKRLGPRPSLLPTGDPANASQPKPLDLKQKRAAALPPIQVTKVHEPPRED 988
DB 148 SSGLSQGPARYHPHPPLEFPSPQPDSTPRQ-----EAEFHPSTVTPGYHAPMEPP 201

QY 989 AAPTKEAPP-APPPQNLOPE-----SDAPQQPG-----SSPRGKSRSPAPPADKE 1033
DB 202 TSMFOAPPAGAPPHPHPPQYPGTGVLGGPPWPKGGGAASSVGGPNKGKQHPPTTPTIS 261

QY 1034 AFABEAQKPGDPCWTSGLFPVPPREVIVAS--PHAPDPSASYPAPGHPPLGLHDT 1091
DB 262 VSSGA-----SGAPPTKPTPTTPVGGNLPSPAPPANFPHVTPNLP----- 302

QY 1092 ARPVLPRPTTISNPPPLSSAKHPSVLERQIGALISQMSVOLHVPYSEHAKAPGVPVTMG 1151
DB 303 -----PPALRPLNNAASAPGLCA-----QPLGHLF-SPHANGQ-----GNG 340

QY 1152 LPLMPDKK---LAPFGVKQEQLSPRQAGPPESLGYPTAQEASVLRGTALGSPVGGSI 1208
DB 341 -GLPPGPEKGTLPAP---SPHSLPPASSAPAPMPRFYSSSS--SSAAAASSSSSSSS 393

QY 1209 TKGIPSTRVPSDAITYRGSITHTGTPADVLYKGTITRIIGDPSRLDRGREDSLPKGHV 1268
DB 394 SSASP---FPASQALP---SYPHSPFP---TSLSVSNQPPKY---TQPSLPSQAV 437

QY 1269 IYEGKGHVLSEGGMSVTQCKEDGRSSGPPHETAAKPTYDMWEGRVGRAISSASIE 1328
DB 438 -----WSQGGP-----PPPPY-----GRLLNNAH 458

QY 1329 -----GLMGRAIPPRHSPHPLKEQHHRIGSITQIPRSYVEAQEDYLRRRAKLKR 1380
DB 459 PGPPFPSTGAQSTAHPPV--STHH-----HH-----HOCQQQQQQQQQQH 498

QY 1381 EGTTPPPPPSRDLTEAYKTQALGPLKPAHEGLVATVKEAGRSIHEIPRE-----ELR 1434
DB 499 HGNSGPPPP-----GAPPHLEGSSHHAPHYAMSPSLGSLR 535

QY 1435 HTPLEPL-APRLKEGSITQ-----GTPLKYDTGASTTGSK-----KHDVRSI--IGSPGR 1482
DB 536 PYPPGPAHLPPPHQVSQAGPNPPVSSNSNSSTSSQGSYPCSHFSPSQGQAP-Y 594

QY 1483 TFPPVHPLVDNADARALERACVEESLKSRRPGTASSGGSIARGAPVIVPELKGKQPSPLT 1542

Db 595 PFPFV-PTVTSSA-----TLSTVIATVASSPAGYKTASPPGPPYCKRAPSPCA 643
QY 1543 YEDGAPFAGHLPRGSPVTMEPTPRLQEGSLSSSKASQDRKLTSTPREIAK-SHSTVP 1601
Db 644 YKTATPP--GYKP-GSPSPFTGTTPGYRG-----TSPGAPGTFFKGS--P 685
QY 1602 EHHPHIPSPYBHLIRGVSGVDLYRSHIPLAIFDPTTIPRGIPLDAAAAYLPRHLAPNTY 1661
Db 686 TVGPGPL-----PPAGPSGLP-----SLPPPA 709
QY 1662 PHLYPPYLIRGYPDTAALENQTIINDYITQQMHNTATATAQADMLRGLSPRESSLA 1721
Db 710 PASGPP-----LSATQIKQEPABEYET----- 731
QY 1722 LNYAAGRGIIDLSQVPHLPVLPPTGTPATAMDRLAYLPTAPOPFSS--RHSSSPLSPG 1780
Db 732 -----PESVPPPARSPSP--PKVDVPSHQSARFNKHLDRGFNSCARSDLYFVPL 782
QY 1781 GPTHLT-----PTTSSSERDRDRDREREKESILTSTTVE--HA 1824
Db 783 EGSKLAKRADLVKVRREAEQRAEKEKEREKEREKEREKEREKEREKEREKEREKEREK 842
QY 1825 PIWRPGTEQSSG 1884
Db 843 PVECP-----SLGVPVPH-----RPPFPGS--AVATVPPYL-GDTP 876
QY 1885 GIITAVBSPKTVLSTSTSSPVPAATFFPATHCPGLG-----GTLGDIVVPTLMEPVLLPK 1940
Db 877 ALRTLSEYARPHVMSVGNRHPF-----YVPLGAVDPGLLGYNNVPALYSSDDPAAR 926
QY 1941 EAPVAPPERPRADTGHAFKAPPARSGLEPASS--PSKSGSPRPLVPVPSGHATTIARTPA 1999
Db 927 EREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKEREKERE 980
QY 2000 KNLAHPHASPP-----PAPPASADPHREKTQSKPFSIQELESRLSGVHGSSYSPE 2051
Db 981 LHPFPFPLSGPLERERLALAAAGPALRPD-----MSYAE-RLAAERQHAERVAAL 1029
QY 2052 GVEVPSVSSPSLTHDKLPHLELDKSHLEGELPKQPGVKLGGEAAHLPHLRPLPE 2111
Db 1030 GNDPLARLQMLNVT-----PHHGH-----SHTHSHLHLHQDAIHAASASVH-FLIDPL-- 1078
QY 2112 SQSSSPLQAPGVKGQRVVTIAQ-----HISEVITQDYTRHH-----POOL 2155
Db 1079 ---ASGSHLTPYPAG-----TLPNPLPHPLHENEVL-----RHQLFAAPYRDLPASL 1125
QY 2156 SAPLAPLYSPGASCPVLDLRRPSPD-----LYLPPPDHGA--PARGSPHSGGKRSPEP 2209
Db 1126 SAPMSA-AHQAMHAQSAELQRLALEQQQWLHAHPLHLSVPLPAQEDYYSHLKKESDKP 1184

RESULT 45
BCL9 HUMAN STANDARD; PRT; 1426 AA.
ID BCL9 HUMAN AC 000512;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE B-cell lymphoma 9 protein (Bcl-9) (Legless homolog).
GN BCL9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC Tissue=Fetal brain;
RX MEDLINE=98158621; PubMed=9490669;
RA Willis T.G., Zaiberg I.R., Coignet L.J.A., Wlodarska I., Stul M.,
RA Jadayel D.M., Bastard C., Treleaven J.G., Catovsky D., Silva M.L.M.,
RA Dyer M.J.S.;
RT "Molecular cloning of translocation t(1;14)(q21;q32) defines a novel

Db 1009 MMSRMSKAMPSSSTLYHDAIKTVASSDDDDSPARS--PNLPSMNMNP-----GMG 1057
QY 2043 YHGSSYEGVEPVSPVSLTHDKLPKHLKLELDKSHLEGELAPKQPGVVKLGGEAAH 2102
Db 1058 INTQNPRISSGNPVVPM--PTLS-FMGMTQPL-----SHSN-----QMPSPNAVQ----- 1099
QY 2103 LPHRLPLPESQSSSPLLOTPAGVKGHQVTVTLAQHISEVITQDYTRHHPOOLSAPLPAP 2162
Db 1100 -----PNIPPHGVPM--GFGLMHNPIIM-----GHGSQEPMPVQ 1133
QY 2163 LYSPPGASCPVLDLRRPPSDLYLPPDHGAPARGSPHSEGG-----KESPEPN 2210
Db 1134 RMGFPQGGPPV--QSPQOQVFPF---HNGESGGGSPGPGMGPGEGPLGRPSNLPSS 1187
QY 2211 KTSVL--GGGSDGIEPVSPGEMTBPGRHSRAVYPLLYRDGEQTBPGRMSKSPGNTSQP 2268
Db 1188 ADAALCKPGGPG-----PDSFTVLGNSMPSVF-----TDPDLQEVIRPGATGIP 1232
QY 2269 PAFSFKLTESNAMYKSKQKOEINKLNTNHNNEPEYNIISQGTETEFNNPAITGTGLMTYR 2328
Db 1233 EFDLSRIIPSEK--PSQTLQVFRGEVPGRKQPO---GPGGPFHMQGMG----- 1278
QY 2329 SQAVQEHASTWGLEAIRKALMGKYDQWESPPLSANAFNPLNASLSLPAAMPIT----- 2384
Db 1279 -----EOAPRMGL-----ALPGM-----GGPGVGTPTDPLGTAPSMPCGNMRPPAF 1321
QY 2385 AADGR--SDHTLTSPGGGKAKVGRSPSRKAKSPAGLAG--DRPPSVSSVHSEGCNR 2441
Db 1322 LQCGMGMGHHRMSP--AQSTMFGQPTLMNPAAGVGMIFGKDRGPAGLYTH----- 1371
QY 2442 RTPLTNRVWEDRPSAGSTPPFYNPLIMLQAGVNAS-----PP--PPGLPAGSG 2489
Db 1372 -----PGPVGS-----PGMWSMQGMGPGQNMIMPPQMRPGMAADVG 1410

RESULT 46
GSRI HUMAN STANDARD; PRT; 1509 AA.
AC GSNZM4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glioma tumor suppressor candidate region gene 1 protein.
GN GLTSCR1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=20175430; PubMed=10708517;
RA Smith J.S., Tachibana I., Pohl U., Lee H.K., Thanarajasingam U.,
RA Portier B.P., Ueki K., Billings S., Ramaswamy S., Mohrenweiser H.W.,
RA Scheithauer B.W., Louis D.N., Jenkins R.B.;
RT "A transcript map of the chromosome 19q-Arm glioma tumor suppressor
region.";
RL Genomics 64:44-50(2000).
CC -!- TISSUE SPECIFICITY: Expressed at moderate levels in heart, brain,
CC placenta, skeletal muscle, and pancreas, and at lower levels in
CC lung, liver, and kidney.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see [http://www.ebi.ac.uk/ebis-sib.ch/](http://www.ebi.ac.uk/ebis-sib.ch/announcement/)
CC or send an email to license@sib.ac.uk).
CC -----
CC EMBL; AF182077; AAF62874.1; --
DR Genew; HGNC:4332; GLTSCR1.
DR MIN; 605690; --
FT DOMAIN 37 POLY-GLY.

FT DOMAIN 884 889 POLY-PRO.
FT DOMAIN 1214 1225 POLY-SER.
FT DOMAIN 1282 1286 POLY-PRO.
FT DOMAIN 1294 1304 POLY-PRO.
SQ SEQUENCE 1509 AA; 152991 MW; 7C5144F443CE6821 CRC64;
Query Match 2.7%; Score 360; DB 1; Length 1509;
Best Local Similarity 23.0%; Pred. No. 6.4e-05;
Matches 249; Conservative 104; Mismatches 420; Indels 308; Gaps 51;
QY 745 SSDTESIPSPHTEAAKDTGQNGPKPPATLGDAGPPG---PPTPRRTSRA-----PIBPT 797
Db 514 SLPTQSQAPAGPAATTVLQGVLPSSAVAMLNTDGLVQVATPAAATGEAAPVITVQPA 573
QY 798 P-ASEATGAPTP-----PPAPPSP-SAPPVVPKEEKEEETAAAPVVEEGEOKP 845
Db 574 PQAPPAVSTPLGLQOQPAQPPQAAPQAATTPQSPGLASSP--EKIVLQGP 631
QY 846 PAABELAVDTG-----KAEFPVKSECTEAEAEAGPAKGDAAEABATAG 889
Db 632 PSATPTAILTQDSLQMFLPQERSQOPLSAEGPHLSVPASVIVSAPPQAQDPAPATPAVK 691
QY 890 ALKAEKEGSGRATTAKSSCAPQSDSSA-----TCSADEVDEAEAGDKNLLSPR 941
Db 692 A-----GLGQPADSQASAPAPQIPAAAPLAKGPGSSPSLPHQAPLGDSPHLPSPH 744
QY 942 PSL--LPTGTGPRANASP--OKPLDLKQLKORAAAIPIQVTKVH--EPPREDAATKP 994
Db 745 PTRPSPRPSPQSVSRPPSPPLHPCPPQAPPTLPGIFVIONQLGVPPPASNPAPTAP 804
QY 995 APPAPP--PPQNLQESDAPQOP-----GSSPRGKSRSPAP-PADKEAFAAEAAQ- 1040
Db 805 GPPQPLRPQSPGEGPLPAPPLPPSPSTSSAVASSSETSSRLPAPTSDFQLOFPSPSQ 864
QY 1041 --KLPGDPPCW-----TSGLPFPVPPRE--VIKASPHAPDPSA-----FSYAPPGHPLPLG 1087
Db 865 PHKSFTPTPTLHLVPEAPAAPPPPTFTFQVMTTTPPALPQKALLERHQVPSG---II 920
QY 1088 LHDTRAPVLRPPPTISNPPPLISSAKHPSVLEROIGALSQMSVQLHVPYSEHAKAPVP 1147
Db 921 LQNKAGAPAAPQTSLSGLTSPA--ASVL-----VSGQAPSGTPTAP--SHAAPAPM 971
QY 1148 VTMLGP--LPMDPKLAFFSGVQEQSLSPROQAGPPESLGVPTAQEASVLRTALGSVPG 1205
Db 972 AATGLPPLPAENKAF--SNLPTLNKAKAASSGPGKPSGLQYESKLSGLK- 1021
QY 1206 GSITKGPSTRVPSDSAL-----TYRGSITH-----GTPA--DVL-----YKGI- 1243
Db 1022 -----PPTLQPSKEACFLEHLKHQGSVLHPDYKTAFFSFEDALHRLPHYVYQAGALP 1074
QY 1244 -----TRIIGED-----SPS----- 1253
Db 1075 SPDSYHKVDEBEFETVSTQLKRTQAMLNKYLLLEBSRRVSPSAEMWIDRMFTQSEKT 1134
QY 1254 --RLDRGREDSLPKGHV-----IYEGKKGHVLSYEGGMSVTQCSKEBGRSSSGPPH 1302
Db 1135 TLALDKLQAKKEKPDYVSSSSSLGLPIAASSEGHRLPGHGLS---SSAPGASTQPPPH 1190
QY 1303 -----ETAAPKRTYDMMEGRVGRALISSASIEGLMGRALPPEKHSPLHKEQHI 1351
Db 1191 LPTKLVIHGGAGGSPSVTW-----ARASSLSLSSSSSSSSAASLSDADSDGMPSPNRPP1 1246
QY 1352 RGSITQIPRSYVAQEDY-----LRREAKL-----LKREGTPPPPPSRDITAEVK 1398
Db 1347 -----KTY-BARSIGLKLKQKAGLSKVVNHTALDPVHQPPPPATLKVKEP- 1294
QY 1399 TQALGPLKLPAHEGLVATVKEAGRSIHEIPREBLR-----HTPELPL----- 1441
Db 1295 -----PPRPPPPPP-PTGQMNGTVDHPHPAAEPKPLGTATPHCPRLPLAKTYRENVGG 1346
QY 1442 --APRPLKBSITOGT--PLKYDTGASTTSGKKHDVRSLLIGSPGRTFPFPHPLDVMADAR 1497
Db 1347 PGAPEGTAGARGGSPAPLPKAVDEATSG-----LIRELA 1382

QY 1498 ALERACYBESLKRSP-----GTASSSGGSIARGAPVIVPELCKPROSPILTYEDHGAFFACH 1553
 Db 1393 AVEEDLYQMLKGPPEPAASAAQTQDPDWEAPGLPPAKRKSESP---DVDQASFSDD 1439
 QY 1554 LPRGSPVMTREPTPLRQBSLSSSSKASQDRKLITSTPREI-AKSPHSTVPEHHPHPIPSYE 1612
 Db 1440 SPQDD--TLTEHLQSAIDSLNLOQAPGRTAPSPYPAASAGTPASPLHRPEAYPSS 1497
 QY 1613 H 1613
 Db 1498 H 1498

RESULT 47
 NEST_HUMAN
 ID NEST_HUMAN STANDARD; PRT; 1618 AA.
 AC P46681; O00552;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Nestin.
 GN NES.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RX MEDLINE=93123384; PubMed=1478958;
 RA Dahlstrand J., McKay R.D.G., Zimmerman L.B., Lendahl U.;
 RT "Characterization of the human nestin gene reveals a close
 evolutionary relationship to neurofilaments.";
 RL J. Cell Sci. 103:589-597(1992).
 RN [2]
 RP SEQUENCE OF 296-309 FROM N.A.
 RX MEDLINE=99117355; PubMed=9917366;
 RA Yaworsky P.J., Kappen C.;
 RT "Heterogeneity of neural progenitor cells revealed by enhancers in
 the nestin gene.";
 RL Dev. Biol. 205:309-321(1999).
 CC -!- TISSUE SPECIFICITY: CNS STEM CELLS.
 CC -!- DEVELOPMENTAL STAGE: UPON TERMINAL NEURAL DIFFERENTIATION, NESTIN
 IS DOWN-REGULATED AND REPLACED BY NEUROFILAMENTS.
 CC -!- SIMILARITY: Belongs to the intermediate filament family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; X65964; CRA46780.1; -;
 DR EMBL; AF004335; AAB64426.1; -;
 DR PIR; S21424; S21424.
 DR Genew; HGNC:7756; NES.
 DR MIM; 600915; -;
 DR GO; 0005882; C:intermediate filament; NAS.
 DR GO; 0007417; P:central nervous system development; NAS.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; filament; 1.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neuron.
 FT DOMAIN 1 7 HEAD.
 FT DOMAIN 8 312 ROD.
 FT DOMAIN 313 1618 TAIL.
 FT DOMAIN 8 43 COIL 1A.
 FT DOMAIN 44 55 COIL 1.
 FT DOMAIN 56 151 COIL 1B.
 FT DOMAIN 152 172 LINKER 12.

FT DOMAIN 173 191 COIL 2A.
 FT DOMAIN 192 194 LINKER 2.
 FT DOMAIN 195 312 COIL 2B.
 SQ SEQUENCE 1618 AA; 176704 MW; C9E9AA48C66534D0 CRC64;
 Query Match 2.7%; Score 358; DB 1; Length 1618;
 Best Local Similarity 21.1%; Pred. No. 7.9e-05;
 Matches 323; Conservative 196; Mismatches 511; Indels 502; Gaps 77;
 QY 97 KSEMFIESKRPL--ELLP--DPLLRPSPL---LATQPA--GSEDLTKDRSLTGKLEP 147
 Db 329 KLELQFPRTPEGRRLGSLLPVLSPTLSPPLATLETPVPAFLKNQFLQARTTFLASTP 388
 QY 148 VSPSPPHTPPELVPRLSKEELIQMDRVDREITMVSQOISKL-----KKQOQLE 201
 Db 389 I-PPTQAPSP-----AVDAETRAQDAPLSLLOTTQGRKQAPFLR 428
 QY 202 EEA--AKP-----VSP-----PIESKH----- 223
 Db 429 AEARVAIPASVLPCEPFGQORQEAQSPEDHASLAPLSPDHSSLEAKDGSGGSRV 488
 QY 224 -----RSLVQIYDE---NRKAEAAH-----RILEGL 248
 Db 489 FSICRGEQIMGLVEKETAIGKVSVSLQOEIWEEDLNKKEIQDSQVPLEKETLKS 548
 QY 249 GPQVE---LPLYNQPSDTROYHENIKINQAMRKLLILYFKERNHARKQWKQKFCQRYDQL 305
 Db 549 GEEIQESLKTLENQSHETLE-RENQECPRSLBEDL-----ET 584
 QY 306 MEALEKKVERI-----ENRRRAKESKVREYEQFPEIRKQELQERMSRVQORG 358
 Db 585 LKSLKENKAIKCGGSETSRKGCQOLKPTKEDTQTLQ-SLOKENQELMSLEGNLE 643
 QY 359 SGLSMSARSEHEVSEIIDGLSEQENLEKQMRQAVIPPMYLDADQORIKFINNGLMAD 418
 Db 644 TFL-FPGTENQELVSSLOENLESITALEKENQPLRSPEV---GDEALRPLTKEN--QE 697
 QY 419 PMKVYKDRQVMNWMWSEQEKETER--EKFMQHPKNFGLIASFLERKTVAECVLYVYLTKN 476
 Db 698 PLRSLED-----ENKEAFRSLEKNOEP-----LKTLEEDQSIVRPLET 737
 QY 477 ENYKSL-----VRRSYRRGKSQQQQQQQQQQQQQQQMPRSSQEEKDEKEKEAE 530
 Db 738 ENHKLSRLEEQDETTLTLEKETQRRRSILGSDQMTLRP-----PEKVDLEPLKSLD 791
 QY 531 KEEKEPVENDKEDLLIKEKTDITSGEDNDEKAVASKRKTANSQGRKRGRITRSMANEA 590
 Db 792 QETARP-LENENOEFKSLKEES-----VEAVKSLTEILES-----LKSAGQE 834
 QY 591 NSEEAITPQQAELASLMELNESSRWTEEN-----ETAKKG-----LLEH 630
 Db 835 NLETLKSPETQAPL-----WTPPEINKSGNNESSRKGNSTRITGVCSPRDIQTP 884
 QY 631 GRNWSAIARMVSGKTVSQ-----KNFYNYKKQNLEILQOHKLKMKERNARRKK 683
 Db 885 GRGESGIIRISGSMPEGFEISRGVDKESQSNLEEBENLGKGYQESL-----RSLBEEG 939
 QY 684 KKAFAAEEAFAFPVVEDEME-----ASGVSGNEEENVEEAALHAGSNE--VPRGE 735
 Db 940 QELPQSAVQVQWEDTVEKDQELAQESPPGMAVENKDEAELNREOQGTGKEVEVEQGE 999
 QY 736 CSGPATVNNSSDTEIPSP---HTEAAKDTGQNGKPPATL---GADGPPPPPTPTPTS 790
 Db 1000 LNA-----TEEVWFPGEGHPENPEKQRLGVEGASVKGAGLQ----- 1039
 QY 791 RAPIEPTPASEATGATPPPPAPSPGAPPPVVPKKEKEEETAAPVVEEKEOKPPAAEE 850
 Db 1040 ----DPEGQQVGTGQLAQPGQLPEALIEPLVEDDVAQGDQASPEVMLGSE---PAMGE 1092
 QY 851 LAV-----DTGK-----AEPVSEKTE-----BAEERPAKGDAAEAET 886
 Db 1093 SAAGAEPGLQGVGLGDPGLHLTREVMPEPPEESLEAKRVQGLEGR---KOLEEA--- 1147

DT 01-MAY-1992 (Rel. 22, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(V) chain precursor.
GN COL5A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 556-565.
RX MEDLINE=91302336; PubMed=20711595;
RA Takahara K., Seto Y., Okasawa K., Okamoto N., Noda A., Yaoi Y.,
RA Kato I.;
RT "Complete primary structure of human collagen alpha 1 (V) chain.";
RL J. Biol. Chem. 266:13124-13129(1991).
RN [2]
RP SEQUENCE OF 621-822.
RC TISSUE=Chorioamniotic membrane;
RX MEDLINE=89227189; PubMed=2496661;
RA Sever J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of three
cyanogen bromide-derived peptides from human alpha 1(V) collagen
chain.";
RL Arch. Biochem. Biophys. 271:120-129(1989).
RN [3]
RP SEQUENCE OF 823-950, AND HEPARIN-BINDING.
RX MEDLINE=90366601; PubMed=2203476;
RA Yaoi Y., Hashimoto K., Koitabashi H., Takahara K., Ito M., Kato I.;
RT "Primary structure of the heparin-binding site of type V collagen.";
RL Biochim. Biophys. Acta 1035:139-145(1990).
RN [4]
RP SEQUENCE OF 556-571.
RC TISSUE=Placenta;
RX MEDLINE=92239022; PubMed=1571108;
RA Mann K.;
RT "Isolation of the alpha 3-chain of human type V collagen and
characterization by partial sequencing.";
RL Biol. Chem. Hoppe-Seyler 373:69-75(1992).
RN [5]
RP SEQUENCE OF 565-576; 756-772; 1012-1029; 1219-1232 AND 1465-1477.
RC TISSUE=Chorioamniotic membrane;
RX MEDLINE=94237164; PubMed=8181482;
RA Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champlaud M.F.,
RA Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT "Diversity in the processing events at the N-terminus of type-V
collagen.";
RL Eur. J. Biochem. 221:987-995(1994).
RN [6]
RP DISEASE, AND VARIANT EDS-I SER-1639.
RX MEDLINE=97195540; PubMed=9042913;
RA de Paape A., Nuytincx L., Hauser I., Anton-Lamprecht I.,
RA Naeyaert J.M.;
RT "Mutations in the COL5A1 gene are causal in the Ehlers-Danlos
syndromes I and II.";
RL Am. J. Hum. Genet. 60:547-554(1997).
RN [7]
RP VARIANTS EDS-1 SER-530 AND ASP-1489.
RX MEDLINE=20068401; PubMed=10602121;
RA Giunta C., Steinmann B.;
RT "Compound heterozygosity for a disease-causing G1489D and
disease-modifying G30S substitution in COL5A1 of a patient with the
classical type of Ehlers-Danlos syndrome: an explanation of
intrafamilial variability?";
RL Am. J. Med. Genet. 90:72-79(2000).
CC -!- FUNCTION: Type V collagen is a member of group I collagen
component of nearly ubiquitous distribution. Type V collagen binds
to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
CC -!- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
most tissues and trimers of one alpha 1(V), one alpha 2(V), and
one alpha 3(V) chains in placenta.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.

CC -!- PTM: Sulfated on 40% of tyrosines.
CC -!- DISEASE: Defects in COL5A1 are a cause of Ehlers-Danlos syndrome
type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
gravis. EDS-I is a connective-tissue disorder characterized by
loose-jointedness and fragile, velvety, stretchable, bruisable
skin that heals with peculiar 'cigarette-paper' scars. Inheritance
is autosomal dominant.
CC -!- DISEASE: Defects in COL5A1 are a cause of Ehlers-Danlos syndrome
type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
syndrome mitis. Inheritance is autosomal dominant.
CC -!- SIMILARITY: HIGH, TO ALPHA 3(V) AND ALPHA 1(XI) CHAINS.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announcement/)
or send an email to license@isb-sib.ch).
CC EMBL; D90279; BAA14323.1; -;
DR PDB; 1A89; 18-NOV-98.
DR PDB; 1A9A; 18-NOV-98.
DR Genew; HGNC:2209; COL5A1.
DR MIM; 120215; -;
DR MIM; 130000; -;
DR MIM; 130010; -;
DR GO; GO:0005598; C:collagen type V; TAS.
DR InterPro; IPR008161; Clg_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR008985; ConA_like lec.gl.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR001791; laminin_G.
DR InterPro; IPR003129; TSPN.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF02210; TSPN; 1.
DR ProDom; PD000007; Clg_helix; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00282; LamG; 1.
DR SMART; SM00210; TSPN; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Collagen; Signal; Heparin-binding; Sulfation; Disease mutation;
KW Ehlers-Danlos syndrome; 3D-structure.
FT SIGNAL 1 37
FT CHAIN 38 1605
FT DOMAIN 39 230
FT DOMAIN 231 443
FT DOMAIN 444 558
FT DOMAIN 559 1570
FT DOMAIN 1571 1605
FT PROPEP 1606 1838
FT MOD_RES 234 234
FT MOD_RES 236 236
FT MOD_RES 240 240
FT MOD_RES 262 262
FT MOD_RES 263 263
FT MOD_RES 338 338
FT MOD_RES 340 340
FT MOD_RES 346 346
FT MOD_RES 347 347
FT MOD_RES 416 416
FT MOD_RES 417 417
FT MOD_RES 420 420
FT MOD_RES 421 421
FT MOD_RES 570 570
FT MOD_RES 576 576
FT MOD_RES 621 621
FT MOD_RES 627 627
FT MOD_RES 639 639
FT MOD_RES 642 642
FT COLLAGEN ALPHA 1(V) CHAIN.
FT TSP N-TERMINAL.
FT NONHELICAL REGION.
FT INTERRUPTED COLLAGENOUS REGION.
FT TRIPLE-HELICAL REGION.
FT NONHELICAL REGION.
FT CARBOXYL-TERMINAL PROPEPTIDE.
FT SULFATION (POTENTIAL).
FT SULFATION (POTENTIAL).
FT SULFATION (POTENTIAL).
FT SULFATION (POTENTIAL).
FT SULFATION (POTENTIAL).
FT SULFATION (POTENTIAL).
FT SULFATION (POTENTIAL).
FT SULFATION (POTENTIAL).
FT SULFATION (POTENTIAL).
FT HYDROXYLATION.
FT HYDROXYLATION.
FT HYDROXYLATION.
FT HYDROXYLATION.


```
FT MOD_RES 648 HYDROXYLATION.
FT MOD_RES 654 HYDROXYLATION.
FT MOD_RES 657 HYDROXYLATION.
FT MOD_RES 675 HYDROXYLATION.
FT MOD_RES 678 HYDROXYLATION.
FT MOD_RES 680 HYDROXYLATION.
FT MOD_RES 686 HYDROXYLATION.
FT MOD_RES 690 HYDROXYLATION.
FT MOD_RES 696 HYDROXYLATION.
FT MOD_RES 705 HYDROXYLATION.
FT MOD_RES 708 HYDROXYLATION.
FT MOD_RES 717 HYDROXYLATION.
FT MOD_RES 720 HYDROXYLATION.
FT MOD_RES 726 HYDROXYLATION.
FT MOD_RES 732 HYDROXYLATION.
FT MOD_RES 744 HYDROXYLATION.
FT MOD_RES 750 HYDROXYLATION.
FT MOD_RES 756 HYDROXYLATION.
FT MOD_RES 762 HYDROXYLATION.
FT MOD_RES 765 HYDROXYLATION.
FT MOD_RES 771 HYDROXYLATION.
FT MOD_RES 774 HYDROXYLATION.
FT MOD_RES 780 HYDROXYLATION.
FT MOD_RES 789 HYDROXYLATION.
FT MOD_RES 795 HYDROXYLATION.
FT MOD_RES 804 HYDROXYLATION.
FT MOD_RES 807 HYDROXYLATION.
FT MOD_RES 810 HYDROXYLATION.
FT MOD_RES 816 HYDROXYLATION.
FT MOD_RES 819 HYDROXYLATION.
FT MOD_RES 834 HYDROXYLATION.
FT MOD_RES 846 HYDROXYLATION.
FT MOD_RES 861 HYDROXYLATION.
FT MOD_RES 864 HYDROXYLATION.
FT MOD_RES 870 HYDROXYLATION.
FT MOD_RES 873 HYDROXYLATION.
FT MOD_RES 876 HYDROXYLATION.
FT MOD_RES 882 HYDROXYLATION.
FT MOD_RES 888 HYDROXYLATION.
FT MOD_RES 891 HYDROXYLATION.
FT MOD_RES 897 HYDROXYLATION.
FT MOD_RES 903 HYDROXYLATION.
FT MOD_RES 906 HYDROXYLATION.
FT MOD_RES 930 HYDROXYLATION.
FT MOD_RES 945 HYDROXYLATION.
FT MOD_RES 1017 HYDROXYLATION.
FT MOD_RES 1020 HYDROXYLATION.
FT MOD_RES 1023 HYDROXYLATION.
FT MOD_RES 1029 HYDROXYLATION.
FT MOD_RES 1221 HYDROXYLATION.
FT MOD_RES 1224 HYDROXYLATION.

Query Match 2.7%; Score 352; DB 1; Length 1838;
Best Local Similarity 20.5%; Pred. No. 0.00014;
Matches 375; Conservative 104; Mismatches 595; Indels 756; Gaps 90;

Qy 842 EOKPPAAEELAVDTGKAEPEVKSE-CTBEAEEG-----PAKGD 879
Db 238 EHYSPDCDVTAPDTPQSDPNPDEYTYEGDEGETYYEYYPEDPDLEKTEPSK-KP 296
Qy 880 ABAABATAGALKAEKKEGGGRATTAKSGAPQSDSSSATCSADEVDEAGGDKNRLLS 939
Db 297 VEAAKETTE-----VPEELTPTTEAAPMPETSEGACKEEDVG 334
Qy 940 -----PRPSLLTPT-----GDPRANASQPKPLDLKQLKORAAAIPIQVTKVHEPP 985
Db 335 IGDYDYVFESEYTPSPYDLDLTYGEEN--PDQPTD-----PGAGA-----EIP 377
Qy 986 REDAAPTAPAPPPPO-----NLOPE-----SDAQPGSGSRGSRSPAPP 1029
Db 378 TSTADTNSNPAPPPGGADLGEFTTEIRNLNLDENYDYPDYDTSSP--SEIGOMP 435
Qy 1030 ADKEAF-----AAEAQKLPDPPCWTGLPFPFPPREVIKASPHAPDPSAFSAPP 1081
```

```
Db 436 ANODTIYBIGGPRGEKGQK--GEPAITEPGMLIEGPP-----GPEGAGLPGPPG 484
Qy 1082 HPLPLG-----LHDTARPVLP RPPTTISNPPPLI-----SSAKHP--SVLER 1120
Db 485 TWGPTGVQVDGERGPPGRPGLPAGADGLPGPGTGMMLPFRFGGGDAGSKGPMWSAQES 544
Qy 1121 QIGAISQMSVOLHVYPYSEHAKAPVGPVTM--GLPLPMDPKKLAPFS--GVKQE-----QLS 1173
Db 545 QAAILQOARLAL-----RGPAGPMGLTGRPGVGP-----PGSGGLKGEPPGVGPG 592
Qy 1174 PRQAGPPESLGVPTAQEASVLRGTALGSVGGSSITKGI PSTRVPSDSAITYRGSI THGT 1233
Db 593 PRGVQGPFPAGKPCR-----RCRA-----GSDGARGMPGQTGP----- 626
Qy 1234 PADVLYKGTITRIIGEDSPSLDRGRED--SLPKGHVIVYEGKKHVLSEVEGMSVTQCSK 1291
Db 627 -----KG-----DRGFDGLAGLP-----GEKGH----- 644
Qy 1292 EDGRSSGPPHETAAPKRTYDMMEGRVG--RAI--SSASIEGLMGRAIPPERHSPHHLKEQH 1349
Db 645 RGDPGSGPPGPGDDGERGD--DGEVGRGLPGKPGPRGLLPGKPGPPGPGP-----P 696
Qy 1350 HIRGSITQIGIPRSYVEAQEDYLRREAKLLKREGTPPPPPPSRDLTEAYKTOALGLKLP 1409
Db 697 GVTGMDGQPGKGNVGPQ-----GEFPGPGQGN----- 726
Qy 1410 AHEGLVATVKEAGRSIHIEI PREELRHTPELPLAPRLKEGSI TOGTPLKYDTGASTTGSK 1469
Db 727 GAQGLPG-----PQAGIP--PGEKGPL--GKP----- 750
Qy 1470 KHDVRSLTGSPGRTFPVHVPLDVMADARALERACVYESLKRPGTASSSGSIARGAPVI 1529
Db 751 -----GLPGMEGADGPGHP-----GKEGPPGEKG--QGPP-- 780
Qy 1530 VPELGKPRQSPLTYEDHGAPFAGHLPRGSPVTMBREPTPRLOEGSLSSKASQDRKLTSTP 1589
Db 781 -----GPQGPY-----PGPRGVKG-----ADGIRGLKGTK 807
Qy 1590 REIAKSPHSTVPEHHPHPI SPYEHLLRGVSDVLYRSHI PLAFDPTS-----PRGI----- 1641
Db 808 GE-----KGEDGPGFGKDMGKGDGRGETGPGPGEDGPE 843
Qy 1642 -PLDAAAAYLPHRLAPNPTYPHLYPPYLRGYDPTAALENRQTIINDYITSQQMHNHTA 1700
Db 844 GPKRGGPGNGDPGLPGPEKGLGVPL--PGY-----GRQ----- 879
Qy 1701 TAMAQRADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVPTTGTATAMDRLAY 1760
Db 880 -----GPKGSGIF-----PGFPGANGKGR--G 901
Qy 1761 LPTAPQPPSSRSHSSPLSPGPGPHTLTKPTTTSSSERERDRDRDREREKSILTSITT 1820
Db 902 TPGKPGRGQGTGTPRGGERPGIT----- 927
Qy 1821 VEHAPIWRPTEQSSGSSGSSG--GGGSSSRPASHAHQHSPTSRTQDALQORPSVL 1878
Db 928 -----GKPGKNSGGDPAGPGERGPNCPQGTGPGKPGPPGKGLPGHPQOR 981
Qy 1879 HNTQMKGIITAVEPSKPTVLRSTSTSSPVRPAATFPFATHCPGLGTLDGVTPTLMPEVLL 1938
Db 982 GETGFOG-----KTGPPGPGVVGVPQ-----PTGET----- 1008
Qy 1939 PKEAPRVARPERPRADTGHAF LAKPPARSGLF-----PASSPSKSGSEP----- 1981
Db 1009 -----GPMGERGHGPPGPPGEGQGLPGLAGKEGTGKDGCPAGLPGKDPGPPGLRG 1057
Qy 1982 ---RPLVPVSGHATARTPAKNLAPHASPDPPAP-----PASASD-----PHREK 2025
Db 1058 FPGDRGLPGPVG-----ALGLKNGEPPGPPGAGSGPGRGAGAGPIGIPGPPG 1108
Qy 2026 TQSKPFSIQEILELRSLGSHGSSYEPGVEPVSVSSPSLTHDKGLPKHL-----EELD 2078
```

Db 1109 PQGPP-----GPAGEKGPCKGPGQAGRDGLQGVGLPGPAGVPGPGEDGD 1157
Qy 2079 KSHL-----EGELRPQKPGVKGLEAAHLPHLRPLPESQPSRSSPLLQTARGVK 2127
Db 1158 KGEIGEPQKSGKDGKEGPG--PGTGPQG-----PTGQPGPSGA---DGEQGP 1203
Qy 2128 GQRVVVTLAHTISEVITDYTRHHPQQLSAPLAPLYSPFPGASCPLDLRPPPSDL----- 2183
Db 1204 GQGL--FGQKGE-----GPRGFP--GPGFVGLQGLPG-----PPGKGTG 1243
Qy 2184 ----YLPDPHAPAGSHSEGGKSPENKTVLGGEDGIBVSPPEGMTPEGHSRA 2240
Db 1244 DVQGMGP-----GPPGPRGPGCAP-----GADG--PQPGPGGIGNPG-----A 1280
Qy 2241 VYPLLYRGEOTEPERSMSKSGPNTSQPAPFSKLTESNAVSKQEIKNLTHNRN 2300
Db 1281 V-----GEKEPGEAGEPGSGSGPP-----GPKGERGEK-----GES 1314
Qy 2301 EPEYNISQGTBIFNMPAITGTGLMYRSQAVQEHASTNMGLEAIRKALMGKY----DQ 2356
Db 1315 GPSGAAGPGPK--GPPGDGP-----KSGFGPVGFGDP 1347
Qy 2357 WEESPPLSAPNPLNASASLPAAMPITAADGRSDHTLTPCGGKGAKVSGRPSRKAKS 2416
Db 1348 GPGGPGAGQDGPDKGD-----DGEPGQT--GSPGTPGPGSPGPKRGPGG 1396
Qy 2417 PA-PGLASGDRPSPVSVHSEGDNRRTPLTNRVWEDRPSSAG--STPPFYNPLIMRLOA 2473
Db 1397 PGPGRQGEK--GAKGEAGLEGPKGTGIPQGAPGKPGDGLRGIPGPGV-----EQ 1449
Qy 2474 GVASPPPPGLPAGSGPLAGPHHAWDEPK 2503
Db 1450 GLPGSGPGDPPGPGWPGPLGLKGDGPK 1479

RESULT 53
CA2B HUMAN
ID CA2B HUMAN STANDARD; PRT; 1736 AA.
AC P13942; Q07751; Q13271; Q13272; Q13273; Q99866; Q9UIP9;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 2(XI) chain precursor.
GN COL11A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96032717; PubMed=7559422;
RA Vuoristo M.M., Pihlajamaa T., Vandenberg P., Prockop D.J.,
RA Ala-Kokko L.;
RT "The human COL11A2 gene structure indicates that the gene has not
RT evolved with the genes for the major fibrillar collagens.";
RL J. Biol. Chem. 270:22873-22881(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Tubby B.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 59-807 FROM N.A.
RC TISSUE=Cartilage;
RX MEDLINE=93314796; PubMed=8325374;
RA Zhidkova N.I., Brewton R.G., Wayne R.;
RT "Molecular cloning of PARP (proline/arginine-rich protein) from human
RT cartilage and subsequent demonstration that PARP is a fragment of the
RT NH2-terminal domain of the collagen alpha 2(XI) chain.";
RL FEBS Lett. 326:25-28(1993).
RN [4]
RP SEQUENCE OF 730-1690 FROM N.A.
RX MEDLINE=89340485; PubMed=2760050;
RA Kimura T., Cheah K.S.E., Chan S.D.H., Lui V.C.H., Mattei M.-G.,

van der Rest M., Ono K., Solomon E., Ninomiya Y., Olsen B.R.;
RT "The human alpha 2(XI) collagen (COL11A2) chain. Molecular cloning of
RT cDNA and genomic DNA reveals characteristics of a fibrillar collagen
RT with differences in genomic organization.";
RL J. Biol. Chem. 264:13910-13916(1989).
RN [5]
RP SEQUENCE OF 1-537 FROM N.A.
RX MEDLINE=96435918; PubMed=8839804;
RA Lui V.C., Ng L.U., Sat E.W., Cheah K.S.;
RT "The human alpha 2(XI) collagen gene (COL11A2): completion of coding
RT information, identification of the promoter sequence, and precise
RT localization within the major histocompatibility complex reveal
RT overlap with the K55 gene.";
RL Genomics 32:401-412(1996).
RN [6]
RP ALTERNATIVE SPLICING.
RX MEDLINE=95239468; PubMed=7721876;
RA Zhidkova N.I., Justice S.K., Wayne R.;
RT "Alternative mRNA processing occurs in the variable region of the
RT pro-alpha 1(XI) and pro-alpha 2(XI) collagen chains.";
RL J. Biol. Chem. 270:9486-9493(1995).
RN [7]
RP DISEASE.
RX MEDLINE=20143361; PubMed=10677296;
RA Melkonian M., Brunner H.G., Manouvrier S., Hennekam R.,
RA Superti-Furga A., Kaeerlaeinen H., Pauli R.M., van Esen T.,
RA Warman M.L., Bonaventure J., Miny P., Ala-Kokko L.;
RT "Autosomal recessive disorder otospondylomyelodysplasia is
RT associated with loss-of-function mutations in the COL11A2 gene.";
RL Am. J. Hum. Genet. 66:368-377(2000).
RN [8]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
RN [9]
RP VARIANT OSMED ARG-661.
RX MEDLINE=95163096; PubMed=7859284;
RA Vikkula M., Mariman E.C.M., Lui V.C.H., Zhidkova N.I., Tiller G.E.,
RA Goldring M.B., van Beersum S.E.C., de Waal Malefijt M.C.,
RA van den Hoogen F.H.J., Ropers H.-H., Wayne R., Cheah K.S.E.,
RA Olsen B.R., Warman M.L., Brunner H.G.;
RT "Autosomal dominant and recessive osteochondrodysplasias associated
RT with the COL11A2 locus.";
RL Cell 80:431-437(1995).
RN [10]
RP VARIANTS GLY-593; LYS-824; LEU-879; THR-1316 AND GLN-1600.
RX MEDLINE=98254467; PubMed=9585596;
RA Koga H., Sakou T., Taketomi E., Hayashi K., Numasawa T., Harata S.,
RA Yone K., Matsunaga S., Otterud B., Inoue I., Leppert M.;
RT "Genetic mapping of ossification of the posterior longitudinal
RT ligament of the spine.";
RL Am. J. Hum. Genet. 62:1460-1467(1998).
RN [11]
RP VARIANT W5S GLU-955.
RX MEDLINE=99021942; PubMed=9805126;
RA Pihlajamaa T., Prockop D.J., Faber J., Winterpacht A., Zabel B.,
RA Gledhill A., Wiesbauer P., Spranger J., Ala-Kokko L.;
RT "Heterozygous glycine substitution in the COL11A2 gene in the original
RT patient with the Weissenbacher-Zweymueller syndrome demonstrates its
RT identity with heterozygous OSMED (nonocular Stickler syndrome).";
RL Am. J. Med. Genet. 80:115-120(1998).
RN [12]
RP VARIANT STL3 940-GLY--PRO-948 DEL.
RX MEDLINE=98165506; PubMed=9506662;
RA Sirko-Osadsa D.A., Murray M.A., Scott J.A., Lavery M.A., Warman M.L.,
RA Robin N.H.;
RT "Stickler syndrome without eye involvement is caused by mutations in
RT COL11A2, the gene encoding the alpha-2(XI) chain of type XI
RT collagen.";

Db 692 HPGKEGPGT-----KGNQSPGPGQGLVGYGP-RGVKGVGDLIRGLKHGKEG-EDGPPG 745
Qy 1265 -KGHVIVYEGKKHVLISYEGHSMVTCQKEDG-----RSSGPPHETAAP-----KRTYD 1312
Db 746 FKGDIGVGRGEV-----GVPSRGEDGPEGKGTGTGDPGPPGLMGEGKGLGVP 798
Qy 1313 MMEGRVGR-----AISSASTEGLMGAIPPERHSPHLKEQHHRGSI--- 1355
Db 799 GLPGYVGRQKPGKSLGPFPGCASKGKARGLSGKSGRGERGTPGRCQGRGPRGATGKS 858
Qy 1356 -----TQICPRSYVEAQEDYLRREAKLLKREGTP-----PPPPPSRDUTE 1395
Db 859 GAKGTSGDGHGPPGERGLPQ-----GPNRPGKPGKPGPGPKDGLP 903
Qy 1396 AYKQ-----ALGPLKLPAPHEGLVATVKEAGRSIHEIPRELRHTPELPLAPRPLK 1447
Db 904 GHPGQRBVFGKGTGP-----PGPGVGVPOQAAGET---GPMGERGH---PGPPGPG 952
Qy 1448 EGSITQGTPLKYDTGASTTGSKKHVRSLIGSP-----RTFPVPHPLDVMADAR 1497
Db 953 E-----QGLP-----GTAG--KEGTDGDPGPGAPGKDGAGLGGFPCERGLGTAGGP 999
Qy 1498 ALERACYEESLKRSGTASSG---GSIARGAPVIVPELKGKPRQSPLTIVYEDHGAPFAGHLP 1555
Db 1000 GLKG---NEGSPGPPGAGSPGERGAAGSGPIGP---GRPG-----P 1037
Qy 1556 RGSPTVMEPTPRLOEGLSSSKASQDKLSTPREIAKSPHSTVPEHHPIPSYHLL 1615
Db 1038 QGPPGAAGEKGVPGKEKPIG-----PTGRDVGQGVGL---PGPAGP----- 1076
Qy 1616 RGVSGVDLYRSHIPLAFDPTSPRIGIPLDAAAAYVLPRLAPNPTYPHLYPPYLIRGYD 1675
Db 1077 PGVAGEDGKGEV-----GDPQKGTGKNKGEGHPGPGP-IGPV-----QPG 1120
Qy 1676 TAALENRTIINDYITSOQMHNTATAMAQADMLRGLSPRESSALNAYAGPRTIDLS 1735
Db 1121 AAGADGEPG-----ARGPQGHFG-----AKDEGTRGFNGPFGPIGLQLGPGSG--EKG 1168
Qy 1736 QVPHLPVLVPTGCTPATMDLAVLTPAPOFFSRHSSPLSP-----GGTHLTKP 1798
Db 1169 ETGDVGPMGPPGP-----PGPRFAGNGADGQGGPPGVGNLGGPGEKGP 1215
Qy 1789 TTTSSERDRDRDRDREREKSIITSTTTTVEHAPIWRPGTQSSGSSSGSGG--G 1846
Db 1216 GESGSPGIQGEVPGVKGPRGERGK--ESGQGEFGP---FGAKGPGQDDPKGNPVPVG 1270
Qy 1847 SSSRPASHAHQHSPISPRTQDALQORPSVLHNTGMKIITAVEPSKPTVLRSTSTSSP 1906
Db 1271 FPGDPPGPEG-----GPRGQDQAK-----GDRGEDG-----EPGQP-----GSP 1305
Qy 1907 VRPAATRPPTHCPGLGTLGTVYPLMEPVLLPKAEPRVARPERPRADTGHAFKAPAR 1966
Db 1306 GPTGEPGPG---PLG-----PLG-----PLG-----PLG-----K 1319
Qy 1967 SGLEPASSPSKSGSPRPLVPVPSGHATTIARTPAKNLAPHASDPDPAPPASASDPHREKT 2026
Db 1320 RG--PAGSP--GSEGR-----QGGKAGKDPGALGAPKGTGPVGPAGPAGKPG--DGL 1367
Qy 2027 QSKPFSI-QEELRSLGVHGSYSPEGVPEVPSVSSPSLTHDKGLPKHLELDKSH--LE 2083
Db 1368 RGLGSGVQQGRPGATGQAG-----PPG--PVGPFLGFLRGDAG-----AKGEXGHFGLI 1416
Qy 2084 GELRPKQPGVKLGGEAAHLPHLPLPSQSSPSPLLTQAPGVGHQVRVVTLAQHSI SEVI 2143
Db 1417 GLIGP--PGEQGEKD-----RGLPGQGS-----PQKXGEMGI----- 1448
Qy 2144 TDYTRHHPOQLSAPLAPLAPLYSFFCASCVPVLDLRRPPSDLYLPPPDHGAARGSPHSEGG 2203
Db 1449 -----PGASGPI-----GPGG-----PP--GLPGPAGPKGAKG 1474
Qy 2204 KESPEPNKTSVLGGEDGIEVPSPEGMTEPCHSRSAVYPLLYDDGQGTPEPSRMGSK--- 2260
Db 1475 ATGP-----GGPKGEGVQGPFG--HFGPPGEVIQPLIQMPKKTTRSRVDGSLMQ 1523

Qy 2261 -----SPGNTSQPPAPFSPKLTESNSAMVSKQOEINK 2292
Db 1524 EDEAIPGTGAPGSPGGLIEIFGSL-----DSLREEIQ 1556
RESULT 54
APC_HUMAN
ID APC_HUMAN STANDARD; PRT; 2843 AA.
AC P25054; Q15162; Q15163;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein).
GN APC OR DF2.5.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91335210; PubMed=1651562;
RA Kinzler K.W., Nilbert M.C., Su L.-K., Vogelstein B., Bryan T.M.,
RA Levy D.B., Smith K.J., Preisinger A.C., Hedge P., McKechnie D.,
RA Finnear R., Markham A., Groffen J., Boguski M.S., Altschul S.P.,
RA Horii A., Ando H., Miyoshi Y., Miki Y., Nishisho I., Nakamura Y.,
RA "Identification of FAP locus genes from chromosome 5q21.";
RL Science 253:661-665 (1991).
RN [2]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX TISSUE=Fetal brain;
MEDLINE=91330307; PubMed=1678319;
RA Joslyn G., Carlson M., Thliveris A., Albertsen H., Gelbert L.,
RA Samowitz W., Groden J., Stevens J., Spirio L., Robertson M.,
RA Sargent L., Krapcho K., Wolff E., Burt R., Hughes J.P.,
RA Warrington J., McPherson J.D., Wasmuth J., Le Paslier D.,
RA Abderrahim H., Cohen D., Leppert M., White R.,
RA "Identification of deletion mutations and three new genes at the
RT familial polyposis locus.";
RL Cell 66:601-613 (1991).
RN [3]
RP ASSOCIATION WITH CATENINS.
RX MEDLINE=94082295; PubMed=8259519;
RA Su L.-K., Vogelstein B., Kinzler K.W.;
RT "Association of the APC tumor suppressor protein with catenins.";
RL Science 262:1734-1737 (1993).
RN [4]
RP DISEASE.
RX MEDLINE=95174843; PubMed=7661930;
RA Hamilton S.R., Liu B., Parsons R.E., Papadopoulos N., Jen J.,
RA Powell S.M., Krush A.J., Berk T., Cohen Z., Tetu B., Burger P.C.,
RA Wood P.A., Taqi F., Booker S.V., Petersen G.M., Offerhaus G.J.A.,
RA Tarnette A.C., Giardiello F.M., Vogelstein B., Kinzler K.W.;
RT "The molecular basis of Turcot's syndrome.";
RL New Engl. J. Med. 332:839-847 (1995).
RN [5]
RP DISEASE.
RX MEDLINE=97094176; PubMed=8940264;
RA Eccles D.M., van der Luijt R.B., Breukel C., Bunyan H., Bunyan D.,
RA Fisher A., Barber J., du Boulay C., Primrose J., Burn J., Fodde R.;
RT "Hereditary desmoid disease due to a frameshift mutation at codon 1924
RT of the APC gene.";
RL Am. J. Hum. Genet. 59:1193-1201 (1996).
RN [6]
RP DISEASE.
RX MEDLINE=20243021; PubMed=10782927;
RA Couture J., Mitri A., Lagace R., Smits R., Berk T., Bouchard H.-L.,
RA Fodde R., Almar B., Bapat B.;
RT "A germline mutation at the extreme 3-prime end of the APC gene
RT results in a severe desmoid phenotype and is associated with
RT overexpression of beta-catenin in the desmoid tumor.";
RL Clin. Genet. 57:205-212 (2000).
RN [7]

Db 795 P-----HSENNQDKGLPRPD-----NRDNRLEGNRGNSSSYRGPGQSRM 833
Qy 1392 -----DLTEAVKTOAL--GPLKLPAGEGLVATV-----KEAGRSIHPIPREELRHTPEL 1439
Db 834 ETRDKGLVNRGQAIISRG-----GLVKQEDFRDKWGR--REDSREKNR----- 879
Qy 1440 PLAPRPLKEGSIQTGTPKLYDTGASTTGKXHDV--RSLIGSPGRTFFPVVHPLVDMADARA 1498
Db 880 -----GEGS-----RUGLVRPGSSREKVPGLQSQDR----- 908
Qy 1499 LERACYEESLSRPTASSGSGSIARGAPVIVPELKGPRQSP-----TYEDHGAPFAGHL 1554
Db 909 -----CAAGSRRGPP-----RAGSQERGRLRAGSRERIPRRAGSR 947
Qy 1555 PRGSPVTREPTPLRQEGSLSSKASQDR--KLSTPREIAKSPHSTVPEHPPH--PISPYE 1612
Db 948 ERGPP-----RGQSR--EGLGRSDFRGRGFRPEPGDGKMY---PYHRDEPPRAPWN 999
Qy 1613 HLLRGVGVDLVYRSHIPLAFOPTSIP-----RGIPLDAAAAYILPR 1653
Db 1000 H-----GEE--RGHEEFLDGRNAPMERERLDDWDRERYWRECERDYQDDTLLEYNRED 1051
Qy 1654 HLANPTVPHLYPPVILRGYDPTALENRQTIINDYITSQOMHNN---TATAMAQRADML 1710
Db 1052 RFSAPPSRSHDGD-----RRGPWDDWDRDQMDDEDY--NREMERDMDRDVDRISRPMY 1105
Qy 1711 -----RGLSPRESSALNAAAGPRGIIDLQVPLVLPVPTTGTATAMDR 1757
Db 1106 DRSLDNEWDYGRPLDQESQFR-----DIPSLPLPL--PPLPLDLYRDRD 1155
Qy 1758 LAYLPTAQPFSSRHS--SPLSPGGPTHLTKPTTSSSER-----DRDRERDRD 1808
Db 1156 WR-----EERNREHGYDRDRDGRGELRIEYPERGDTWREKRDVYVDMWDERELSDR 1209
Qy 1809 EREKILSTTTVEHAPTWRCTEQ--SSGSGSGSGGGG---SSSRASHSHAHQSP 1863
Db 1210 WPSVDVDRHSPMAEHMPSHSSSEMMSGSDASLSDQGLGVVLSQR-----QHEII 1261
Qy 1864 SPRTQDALQRPSPVLNTHGKGIITAVEPS-----KPTVLRTSTSSPVRPAATFPATH 1918
Db 1262 LKAQELKMLREKQELQKMDP--GSEFQWADHLPPQESLQNTSS--RCMTYPPGSGY 1317
Qy 1919 CPLGGLGVYFTLMEPVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEPASPFG 1978
Db 1318 RP-----PPMCKPGSIVRP-----SAPPARSSV--PVRTPVP 1350
Qy 1979 SEPRPLVPPV-----SCHATART 1997
Db 1351 IPPPPPPPPPPPPVVIKQTSAVEQERWDEDSFYGLWDTNDEQGLNSEFKSETAAIPSA 1410
Qy 1998 PAKNLAPHASPDPPAPPASADPHREKTQSKPFSIQELRLSLGY--HGSSYSPEGVEPV 2056
Db 1411 PVLPPPPVHSSIPPGPVPMGMP-----MSKPPVQO-----TDVYGHGRDINSTNKVEQI 1461
Qy 2057 -----SPVSSPSL-----TH 2066
Db 1462 PYGERITLRPDPLPERSFTFETHAGQDRYDRERDREPYFDRQSNVIADHRDKRDRETH 1521
Qy 2067 -DKGLPHLELDKSHLEGELRPKQPGVKLGGAHLPLR-----PIPESQSSPILQOT 2122
Db 1522 RDRDRGVIDYDRDRFRDRRRDRDRAQSYRDKKDHSSRRGGRGDFRPSYDRKSDRVPYE 1581
Qy 2123 APGVKGHORVVTLAHISEVITQDYTRHHPOOLSAPLPAPLYSPFGASCPLVLDLRRPPS- 2181
Db 1582 GPSMFGGER-----RTYPE--RWPLPAPLSLHOPPPAPRVE--KKPESK 1622
Qy 2182 ---DLYLDP-----PD-----HGAPRGSHP 2199
Db 1623 NVDDILKPPGRESRPERIVWIMRGLPGSGKTH 1654

RESULT 56
CAL3_MOUSE

ID AC 08121: 061429: 09CEN7: STANDARD; PRT; 1464 AA.
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X DBA: TISSUE=Embryo;
RX MEDLINE=95011609; PubMed=7926795;
RA Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RT complete DNA sequence.";
RL Gene 147:161-168(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Sapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Basak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickinson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smailus D.E.,
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
RT human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogel G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
RT alpha-1 type-III collagen chain.";
RL Gene 61:225-230(1987).
RN [4]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=85131189; PubMed=3972847;
RA Liau G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
RT (III) collagen gene.";
RL J. Biol. Chem. 260:3773-3777(1985).
RN [5]
RP SEQUENCE OF 810-1464 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Embryonic head;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,


```
Dh 713 GPFGASGSLQMPGEGGP---GSPGKGEKGGPGAGAD----- 751
Qy 1599 TVEHPHPHPISEYHLLRQVSDYLRSHIPLAFDPTISPRGIPDADAAAYLPHRLAPN 1658
Dh 752 -----GVPKGD-----GPRG----- 761
Qy 1659 PTYPHLYPPYLRGVPDTAALENRTIINDYITISQMHNTATAMAQRADMLRGLSPRES 1718
Dh 762 PAGP-IGPP-----GPAGQDQKGGGSP--- 784
Qy 1719 SIALNYAGPRGIIDLSQVPHLPVLPPTGPTATAMDRLAYLPTAPQFFSSRSSPLS 1778
Dh 785 --GLPGIAGPRG-----CPG-----ERGEHGGPPGAGF----- 810
Qy 1779 PGGPHLTKPTTTSSEERDRDRDREREKSILSTTVEHAPTWRCGTQSSGSS 1838
Dh 811 PGAPONGEP-----GAKGERGAPGKGGGPPGAGPTGSS-----GPAGPPGQGVKGER 862
Qy 1839 GSSGGG-----GSSSRPASHAHQHSPISRTQDALQQRPSVLHNTGMKGIITAVEPS 1893
Dh 863 GSPGGGTAGTGGGRLCPGPPGNNGNPPGPGSGAPGKGGPPGAGNSG-----SPG 914
Qy 1894 KPTVLRSTSTSPV-----RPAATFPATHCPGLGTLGVYPTLMBPVLLPKPRVARPE 1949
Dh 915 NPGIAGPKGDAGQPGKGPAGGPPGSGPLG--IAGL----- 951
Qy 1950 RPRADTGHAFKAPKPARSGLEPASSPS-----KGSEPRPLVPPVSGHATARTPAKNLAPH 2005
Dh 952 -----TGARGLAGPPGMPG--PRGSPGPGQIGSGKGP-----GASGH 987
Qy 2006 HASPDPPAP-----PASASDPHREKTSQKPFISQELRLSLGVHSGSYSPGVEPVSPV 2059
Dh 988 NGERGPPGPGQGLPGQPGTAGPRDGNPG-----SDGQPRDGSPPGKGDRCEN 1036
Qy 2060 SPSLTHDKGPKHLELDKSHLEGELRPKQPPVKLGGEAAHPLHRLPLPESQFSSP- 2118
Dh 1037 GSPGA-----PGAPGH-----PGPPGVPGSKGDRGETGP---AGPSGAPG 1076
Qy 2119 --LLQTAPE-----GVKGHORVVTLAQHISEVITQDTRHHQQQLSAP 2158
Dh 1077 PAGARGAPGPGPRDQKGETERGSGNGIKGHR----- 1108
Qy 2159 LPAPLYSPFGASCPVLDLRRPPSDLYLPPPDHGAP---ARGSPHSEGGKSPBPNTSV 2214
Dh 1109 -----GPPG-----NPG-----PPGSGAAGHQAIGSPGAPGPRGVPVPHGPP- 1147
Qy 2215 LGGEDGIE-----PVSPP-----EGWTEPHSRSAVYPLLRYDGEQTSPRWGS 2259
Dh 1148 ---GKDGTSGHGPIGPPGPRGNRGERGSEG--SPGH-----PGQPPGPPGPA 1191
Qy 2260 KSP-----GNTS-----QPPAFTSKLTESNAMSVMKSKQKQINKKLANTH- 2297
Dh 1192 PGPCGGGAAAIAGVGKSGSGFPYGGDDPMDPKINTEEINSLKSVNGQIESLISPDG 1251
Qy 2298 NRNEPEYN 2305
Dh 1252 SRKNPARN 1259
```

RESULT 57

```
CA14_MOUSE
ID CA14_MOUSE STANDARD; PRT; 1669 AA.
AC P02463;
DC 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Collagen alpha 1(IV) chain precursor.
GN COL4A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89197932; PubMed=2703490;
RA Muthukumar G., Blumberg B., Kurkinen M.;
RT "The complete primary structure for the alpha 1-chain of mouse
RT collagen IV. Differential evolution of collagen IV domains.";
RL J. Biol. Chem. 264:6310-6317(1989).
RN [2]
RP SEQUENCE OF 1-1154 FROM N.A.
RX MEDLINE=88112221; PubMed=3338568;
RA Wood L., Theriault N., Vogeli G.;
RT "cDNA clones completing the nucleotide and derived amino acid
RT sequence of the alpha 1 chain of basement membrane (type IV) collagen
RT from mouse.";
RL FEBS Lett. 227:5-8(1988).
RN [3]
RP SEQUENCE OF 1149-1424 FROM N.A.
RX MEDLINE=86301886; PubMed=3755692;
RA Nath P., Laurent M., Horn E., Sobel M.E., Zon G., Vogeli G.;
RT "Isolation of an alpha 1 type-IV collagen cDNA clone using a
RT synthetic oligodeoxynucleotide.";
RL Gene 43:301-304(1986).
RN [4]
RP SEQUENCE OF 1276-1669 FROM N.A.
RX MEDLINE=85127033; PubMed=2578961;
RA Oberbaumer I., Laurent M., Schwarz U., Sakurai Y., Yamada Y.,
RA Vogeli G., Voss T., Siebold B., Glanville R.W., Kuhn K.;
RT "Amino acid sequence of the non-collagenous globular domain (NC1) of
RT the alpha 1(IV) chain of basement membrane collagen as derived from
RT complementary DNA.";
RL Eur. J. Biochem. 147:217-224(1985).
RN [5]
RP SEQUENCE OF 1441-1669 FROM N.A.
RX MEDLINE=87250460; PubMed=3597383;
RA Kurkinen M., Condon M.R., Blumberg B., Barlow D., Quinones S.,
RA Saus J., Pihlajaniemi T.;
RT "Extensive homology between the carboxyl-terminal peptides of mouse
RT alpha 1(IV) and alpha 2(IV) collagen.";
RL J. Biol. Chem. 262:8496-8499(1987).
RN [6]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=86196099; PubMed=3009468;
RA Sakurai Y., Sullivan M., Yamada Y.;
RT "Alpha 1 type IV collagen gene evolved differently from fibrillar
RT collagen genes.";
RL J. Biol. Chem. 261:6654-6657(1986).
RN [7]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89066738; PubMed=3198626;
RA Kaytes P., Wood L., Theriault N., Kurkinen M., Vogeli G.;
RT "Head-to-head arrangement of murine type IV collagen genes.";
RL J. Biol. Chem. 263:19274-19277(1988).
RN [8]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=89071759; PubMed=3200851;
RA Burbelo P.D., Martin G.R., Yamada Y.;
RT "Alpha 1(IV) and alpha 2(IV) collagen genes are regulated by a
RT bidirectional promoter and a shared enhancer.";
RL Proc. Natl. Acad. Sci. U.S.A. 85:9679-9682(1988).
RN [9]
RP SEQUENCE OF 1-129 FROM N.A.
RX MEDLINE=88243724; PubMed=3379041;
RA Killen P.D., Burbelo P., Sakurai Y., Yamada Y.;
RT "Structure of the amino-terminal portion of the murine alpha 1(IV)
RT collagen chain and the corresponding region of the gene.";
RL J. Biol. Chem. 263:8706-8709(1988).
RN CC -I- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -I- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure with
CC 2 other chains to generate type IV collagen network.
```

-!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous domain (NC1) at their C-terminus, frequent interruptions of the G-X-Y repeats in the long central triple-helical domain (which may cause flexibility in the triple helix), and a short N-terminal triple-helical 7S domain.

-!- PWM: Prolines at the third position of the tripeptide repeating unit (G-X-Y) are hydroxylated in some or all of the chains.

-!- PWM: Type IV collagens contain numerous cysteine residues which are involved in inter- and intramolecular disulfide bonding. 12 of these, located in the NC1 domain, are conserved in all known type IV collagens.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; J03758; AAA37439.1; -;
 EMBL; M23333; AAA51625.1; -;
 EMBL; J04694; AAA50292.1; -;
 EMBL; X06777; CAA29946.1; -;
 EMBL; X02201; CAA26132.1; -;
 EMBL; M15832; AAA37340.1; -;
 EMBL; M14042; AAA37342.1; -;
 EMBL; M12879; AAA37343.1; -;
 EMBL; M13024; -; NOT_ANNOTATED_CDS.
 EMBL; M13025; -; NOT_ANNOTATED_CDS.
 EMBL; M13026; AAA37344.1; -;
 EMBL; M13027; AAA37345.1; -;
 EMBL; M13043; AAA37346.1; -;
 EMBL; J04448; AAA37437.1; -;
 PIR; A33525; CGMS4B.

MGI; 88454; Col4a1.

GO; GO:0005604; C:basement membrane; IDA.

InterPro; IPR008161; C1g_helix.

InterPro; IPR008160; Collagen.

InterPro; IPR001442; Procollagen4_C.

Pfam; PF01413; C4; 2.

Pfam; PF01391; Collagen; 23.

ProDom; PD000007; C1g_helix; 6.

ProDom; PD003923; ProcollagenC4; 1.

SMART; SM00111; C4; 2.

Extracellular matrix; Connective tissue; Basement membrane;

Repeat; Hydroxylation; Glycoprotein; Collagen; Signal.

SIGNAL 1 27

PROPEP 28 172

CHAIN 173 1669

DOMAIN 173 1440

DOMAIN 1441 1669

DISULFID 1460 1551

DISULFID 1493 1548

DISULFID 1505 1511

DISULFID 1570 1665

DISULFID 1604 1662

DISULFID 1616 1622

CARBOHYD 126 126

CONFLICT 26 186

CONFLICT 186 186

CONFLICT 319 319

CONFLICT 369 369

CONFLICT 403 403

CONFLICT 481 481

CONFLICT 493 493

CONFLICT 712 712

CONFLICT 813 813

CONFLICT 982 982

CONFLICT 1397 1397

SEQUENCE 1669 AA; 160680 MW; 42916B91E52058E9 CRC64;

Query Match

2.6%; Score 348; DB 1; Length 1669;

Best Local Similarity 19.3%; Pred. No. 0.00017;
 Matches 392; Conservative 133; Mismatches 651; Indels 856; Gaps 97;
 QY 733 RGECSGPAT-----VNNSSDTSIP-----SPHTEAAKDTGQNGKPP 770
 Db 28 KDCGSCGCGCHGVKQGERGLPQLQGVGPGMGQEGPHGPGQKGDAGEPL 87
 QY 771 ATLGADGPP-----PGPPTPP-----RRTSRAPTEPTPASATGAP 806
 Db 88 GTKGTGPGGAAGVPGNPLPGIPQDGGPPGPGIPGCGNGTKGERGLPGLPGSGNP 147
 QY 807 TP-----PPAP 812
 Db 148 GPPGLPGMKGDPGEILGHVPTLLKGERGFPPIGMPGSPGLPGLQGVGPPGFTGPP 207
 QY 813 PPSAPPV-----PKEEKEETAAPPVVEEGEQ-----KPPAAEELAVDTG 856
 Db 208 PGPPGPGKEKQMGSSFGPGKDGEGQGVGPPGPPGQAQVKEKGDPAFTKGGKGF 267
 QY 857 KAEFPVKSECTEEAEFGP--AKGDAEAAEATAEGALKAEKKEGSGRATTAKSSGAPQD 914
 Db 268 FPGVPGYGEKGEKPGKQGRGKPGKDGKGERGSPGI-----PGDSGYPLPGRQG- 320
 QY 915 SSSATCSADEVDEAEGDKNRLLSRPSLL--TPTGDPANASPKQLDLKQLKORAAA 972
 Db 321 -----EKGEAG-----LPGPPGTVTGTWPLGKGDGPGAP-----GLRGE 359
 QY 973 IPIQVTKVHEPPREDAAPTKPAPPPQNLQPESDAPQPGS-----1017
 Db 360 -----PKGFPPTGPGPPGPPPTP---GQAGAPFGERGKGGQGFPGVSLP 404
 QY 1018 SPRGKSRSPAPPADKFAAQAQLPGDP--PCWTSGLPFPVPPREVIKASHPADPSAF 1075
 Db 405 GPSGRDGAAGPPG-----PGPPGQPGHTNG-----IIVCCQPGPG- 440
 QY 1076 SYAPPGHPLPLGLHDTARVLPRLPPTISNPPLISSAKHPSVLERQIGAIQSGMSVQLHV 1135
 Db 441 DQGPPTGPGPGL-----TGEVG-----459
 QY 1136 PYSEHAKAPGVMTGLPLPMDPKLAPFSGVKQEQLSPRGQAGPPESLGVPTAQEASVL 1195
 Db 460 -KGQGESCLADTEGL-----RPPPGPQPGGPGGPGQAGKD 499
 QY 1196 RGTALGSPVSGSITKIPSTVPDSAITYRGSITHTGTPADVLKGTITRIIGEDSPRL 1255
 Db 500 RG-----LPGRDGLEGLPGPQ-GSPGLIGPGA--KGEPPGIFPD---MRLKGD----- 542
 QY 1256 DRGREDSLPKHVIYEGKHVLSYEGGMSVTQCSKEDGRSSSGPPHETAAPKRTYDME 1315
 Db 543 -----KGDPPGPGQ-----PGMPGRAGTPGR--DGHP 567
 QY 1316 GRVGRATSSASIEGLMGRAIPPERHSPHLLKEQHHRGSIITQIGIPRSYVEAQEDYLRREA 1375
 Db 568 GLPGPKGSPGSI-GLKGERGPP-----GGV-GFPGS-----596
 QY 1376 KLLKREGTPPPPPSRDLTEAYKTQALGPLKLK-----PAHEGLVATVKEAGRSIHE 1427
 Db 597 ----RGDIGPPPG-----VGPIGVCEKQAGPPGPGSPGLPGPKGEAGKVPL 644
 QY 1428 IPRELHRTPELPLAPPLKSGSITQGTPLKYDTGASTTGSKKHVDVRLSISPGRTFPVPV 1487
 Db 645 PGPPGAAGLPGSPGPGPQGDGRG-PGTPGR-----PGIPGSEK-----GAVGQPGIGFPG 694
 QY 1488 HPLDVMADARALACVVEESLKSPPGTASSG--GSTAR-CAPVIVPELGP-----RQS 1539
 Db 695 -----PGKGVNDGLPGLGRPGSP-----GRPGNGLPGN 724
 QY 1540 PLTYEDHGAFFAGHLP--RGSPTVMTREPTPLQBSLSSSSKASQDRKLTSTPREIAKSPH 1597
 Db 725 PGQOGKGEFGIG-LPGLKQGPGLPGIPGTGKEKSGISGPGVPGEGQGLTGGP----- 776
 QY 1598 STVPEHHPHPTSPYEHLLRGVSGVDLYRSHIPLAFDPTSIIPRGLDAAAAAYLPRHLAP 1657

Db 777 -----LQIRG-----DPG--PPGVQGPAGP----- 795
Qy 1658 NPTYPHLYPPYLI-----RGYPTAALNRQIINDYIITSQQMHH 1697
Db 796 -PGVPGGPGGAMGPPGEGPPGSGGPPGIGKGFPGFGLD----- 837
Qy 1698 NTATAMAQRADMLRGLSPRESSLALNVAAGPRGIIIDLSQVPHLPV--VPPTPGTPATAM 1755
Db 838 -----MPGPKDKSQGLPGLTGOSGL-----PGLPGQQTGPGVFPFG-SK 878
Qy 1756 DRLAYLPAPOPFSSRHSSPLSPGPGPHLTKPTTSSSERERDRDRDRERREKSIL 1815
Db 879 GEMGVMTGPGPGSPGAGTGLPG-----EKG-- 906
Qy 1816 TSTTVEHAPTWRTGTEQSSGSSGSGG--GSSSRPASHAHQHSPIPSPTQDALQQ 1873
Db 907 -----DHG-----LPGSSGFRGDPGKDGKGVGLPGMPSMEHDMG---SMKQKGDQG 954
Qy 1874 RPSVLHNTGKGIITAVEPSKPTVLRSTSTSSPVPAATFPATHCPGLGTLDGVIPTLM 1933
Db 955 EKGQIGPTGDK-----SRGDPGTPG----- 975
Qy 1934 EPLLKPEAPVARPER--PRADTGHAFLAKPPARSGLE-----PASSPSKGE 1980
Db 976 ---VPGKDGQAGHPGPGPGKDRG---LSGTPGSGPLPGKSGVGMGLPGSGPEKG-- 1026
Qy 1981 PRPLVPPVSGHATTARTPAKNLAPHASPDPPASADPHREKTSKPSIQELESLS 2040
Db 1027 ---VPGIPGSGVPGSGEKA-----KGEKGS----- 1052
Qy 2041 LGYHSSYSPRGFPVSPVSSPSLTHDKGL-----PKHLEELDKSHLEGELRPKQPGPV 2094
Db 1053 -----GLPGIGTPGRPGDKDGLAGFPSPGCKEKSAGTPG--MPGSPGR 1099
Qy 2095 KLGEAAHLPHRLPL-ESQSSSPLQTAGVKGHORVTLAHI SEVITQDYTRHPQ 2153
Db 1100 GSPGNIGH-PSGSLPGKDGKGLPGLDGVPGVKG----- 1133
Qy 2154 QLSAPLAPLYSFFCAGFVLDRRPPSDLYLP-----PPDHGAPARGSPHSEGGKRSPE 2208
Db 1134 --EAGLPOT---PGTPGAGKGEPSGD-GIPGAGEKGQGVPGKGFPGFPGSK-GDK 1185
Qy 2209 PNKTSV-----LGG-----GEDGIEPVSPGTEPGHRSRVAVYPLLVRDGE---QTE 2253
Db 1186 GSKGEVGPGLAGSPGIPGVKGEQF--MGPPGQGGPLGTPGHPVEGPKGDRGPGQ 1243
Qy 2254 PSRMGSKPGNTSPPAPFSLKTESNMVSKQKEIN-KKLNTHNRPENYINSQPGTE 2312
Db 1244 PGLFGHPG---MGPPGP-----PGINGPKGDKGNQGW---GAPGVP 1280
Qy 2313 -----IFNMPAITGTLMTYRQVQVQEHASTNMLEAI-----RKALMG-----KYDOW 2357
Db 1281 GPKGDGPGGPGGPGGPGIT-----GSKGDMGLPGVPGFGQKGLPGLQGVKGDQ 1332
Qy 2358 EESPPLSANAFNPLNASLPAAMPITAADGRSDHLLTSPGCGGKAKVSGRPSRKAASP 2417
Db 1333 DQG-----VPGPKGLQGPPGPPGPDYVIK 1357
Qy 2418 APGLASGDRPSSVSHSEGCNRRTPLTNRV-----WEDRPSASGT-PP-PYN 2465
Db 1358 EFGLPFGPGPLGKLGLOQFPFGKQGGVTVGVLPGFPVPGFAGPQKGTGTFPGPG 1417
Qy 2466 PLIMRLQGVNASPPPLGAGSGPLAGP-----HHAWDEEPKPLIC 2507
Db 1418 P-----RGFPFGPGLPMSGMPGTPSVDHGFLVTRHSQTDDP---LC 1460

RESULT 58

KI67 HUMAN

ID_KI67_HUMAN STANDARD; PRT; 3256 AA.

AC P46013;

DT 01-NOV-1995 (Rel. 32, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

15-MAR-2004 (Rel. 43, Last annotation update)
Antigen KI-67.
MKI67.
Homo sapiens (Human).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=94043435; PubMed=8227122;
RA Schlueter C., Duchrow M., Wohlenberg C., Becker M.H.G., Key G.,
RA Flad H.-D., Gerdes J.;
RT "The cell proliferation-associated antigen of antibody Ki-67: a very
RT large, ubiquitous nuclear protein with numerous repeated elements,
RT representing a new kind of cell cycle-maintaining proteins.";
RL J. Cell Biol. 123:513-522(1993).
[2]
SEQUENCE OF 1-31 FROM N.A.
RA Gerdes J.;
RA Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Thought to be required for maintaining cell
CC proliferation.
CC -!- SUBCELLULAR LOCATION: Nuclear. Predominantly localized in the G1
CC phase in the perinuclear region, in the later phases it is also
CC detected throughout the nuclear interior, being predominantly
CC localized in the nuclear matrix. In mitosis, it is present on all
CC chromosomes.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long;
CC IsoId=P46013-1; Sequence=Displayed;
CC Name=Short;
CC IsoId=P46013-2; Sequence=VSP 004298;
CC -!- DEVELOPMENTAL STAGE: Expression of this antigen occurs
CC preferentially during late G1, S, G2 and M phases of the cell
CC cycle, while in cells in G0 phase the antigen cannot be detected.
CC -!- SIMILARITY: Contains 1 FHA domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

EMBL; X65550; CAA46519.1; -;
EMBL; X65551; CAA46520.1; -;
EMBL; X94762; CAA64388.1; -;
PIR; A48666; A48666.
GeneW; HGNC:7107; MKI67.
GK; P46013; -;
MIM; 176741; -;
DR GO; 0008283; P:cell proliferation; TAS.
DR GO; 0000074; P:regulation of cell cycle; TAS.
DR InterPro; IPR00253; FHA.
DR InterPro; IPR008984; SWAD_FHA.
DR Pfam; PF00498; FHA; 1.
DR SMART; SM00240; FHA; 1.
DR PROSITE; PS0006; FHA_DOMAIN; 1.
KW Cell cycle; Antigen; Nuclear protein; ATP-binding; Repeat;
KW Alternative splicing; Polymorphism.
FT DOMAIN 27 76
FT DOMAIN 1000 2928 16 X 122 AA APPROXIMATE REPEATS.
FT REPEAT 1000 1112 1.
FT REPEAT 1122 1234 2.
FT REPEAT 1244 1356 3.
FT REPEAT 1366 1477 4.
FT REPEAT 1487 1598 5.
FT REPEAT 1608 1720 6.
FT REPEAT 1730 1842 7.
FT REPEAT 1851 1964 8.
FT REPEAT 1974 2086 9.

FT	REPEAT	2096	2204	10.	
FT	REPEAT	2214	2326	11.	
FT	REPEAT	2335	2447	12.	
FT	REPEAT	2457	2569	13.	
FT	REPEAT	2579	2689	14.	
FT	REPEAT	2699	2808	15.	
FT	REPEAT	2818	2928	16.	
FT	NP BIND	3034	3041		ATP (POTENTIAL). Missing (in isoform Short). /FtId=VSP_004298.
FT	VARSPLIC	136	495		T -> S (in dbSNP:111106). /FtId=VAR_014858.
FT	VARIANT	3150	3150		K -> E (in dbSNP:8473). /FtId=VAR_014859.
FT	VARIANT	3217	3217		
FT	SEQUENCE	3256	AA; 358741	MW; 578F8C51BED42517	CRC64;
Query Match					2.6%; Score 348; DB 1; Length 3256;
Best Local Similarity					18.7%; Pred. No. 0.00032;
Matches 53;					Conservative 352; Mismatches 1129; Indels 852; Gaps 133;
Qy	4	STQLVAQTWRATEPRYPHSLSYVQIARTHTDVGLEYQHHSRD-YAS			-----51
Db	325	SVQTPSKAVGASFLYEPAKMKTVPQVYQQNS----	PQHKNDLYTTGRRSVNLGKS	380	
Qy	52	-----HLSPGSIIOQRRRPSLL-----SEFQGNERSQELHLRPESHYLP--EL	95		
Db	381	EGFKAGDKTLTPRK-LSTRNRTPAKVEDAADATKPNLSKTRGSIPTDVEVLPTETEI	439		
Qy	96	GKSEMEFIESKRPLELLPDLRLBSPL-----LATQOPAGS-----EDITKORSL	141		
Db	440	HNEPFLWLTQVERKIQKOSLSPEKLTGTAGCMCSGLPSVDINNFGDSINSEGI	499		
Qy	142	TGKLEPVSPSPHPTDPEL--ELVPPR-----LSKEELIQNMDRVREITWVEQOI	190		
Db	500	PLKERRVS--FGGHLRPELDPENLPNTPLKRGAPTCKKSLVWHTPVLLKIIKEQPQP	557		
Qy	191	SKUKKKOQOLEE-----AAKPEPEKVPSPPIESKHSLSLVOIYDENRKAABAAHRI	244		
Db	558	SGKQESGEIHVEVKAQSLVISPPAPSPRKTVPASDQRRRSCKTAPASSSQTEVPKR-	616		
Qy	245	LEGIGPOVELPLYNQPSDTRYHENIKINQAMRKKLILYFKRNHA-----RKWKQK	297		
Db	617	---GGERVATCLQKRVSISSQHDILQ-----MICKSRSGSGASEANLIVAKSWADV	664		
Qy	298	FCQRYDQMLEALEKVERIENNNRRRAKESKVREYEQKPEIRKQELQELRMOSRVCGR	357		
Db	665	V-----KLGNKQOTQVIKHGPR-----SMNKRQRRPATPKPVGUV	702		
Qy	358	SGSLMSAARSEHEVSIIDGLSBOENLEKOMQLAVIPPMPLYDADQORIKIFINMGLMA	417		
Db	703	HSQFSTGHANSP--CTIIIGKAHTEKVHVPARPVVLNNFI---SNQKMDFKEDLSGIA	756		
Qy	418	DPMKVYKDRQVM-----NMWSEQK--ETFREKFMQHPKMFGLIASFLER--	460		
Db	757	EMFPTPVKEQPLTSTCHIAISNSENLIGKQFGQTDGSEEPLLPTSESFGGNVFPFSAQNA	816		
Qy	461	-----KTVAECVLVYLYLTKNENYKSLVRRSYRRRGKSGOQQOQQOQQOQQOQQPMPR	514		
Db	817	AKQPSDKCSASPLPRQCIREHGNVAKTPRNTYKMTSL-	857		
Qy	515	SSQBEKDEKEKEAEKEEKEPEVENDKEDLLKEKTDGSDENDEKEAVASKRGKTANS	574		
Db	858	TSDDTETSPKTVSTVNRSGRSTEFPRNIQKLPVESKSEETNTE---IVECILKRGQKATLL	914		
Qy	575	QGRKRGKITRSMANESEEAITPQQSAELASMEINSSRWTEEMETAKGGLLEHGRNW	634		
Db	915	QORREGEM-----KEIERPFETYK-ENIELKEN-----DEKMKAMKR-----SRT	954		
Qy	635	----SATAFMVGSKTVSOCKNFYNYKKRONLDEILOHKLKMEKENAREKKKAPAAA	690		
Db	955	GQKCAPMSDUTDLKSLPDTE-LMKDTRAGQNLQ--TDHAKAPKSEKG---KITMPQCQS	1009		
Qy	691	SE-EAAPPVPPVVEDEMEAS-GVSGNEBEMVEAEALHASG-----NEVPREGCSGPATVN	743		

Db	1010	LOPEINTPTHTTKQOLKASLGVGVBEELAVGKFTRTSGTTHTRHPADGKSI	TFK	1069
Qy	744	NS-----SDTESI-----	PSPTHEAAKD--TGQNGP	767
Db	1070	ESPKQIILDPAAVRTGMKWPRTPKEAQSLEDLAGKELFQTPGPEESWTDKTTKIAC	1129	
Qy	768	KXPATLGDAGPPPPPTPPRRTSRAPTEP-----TPASEATGAPTPPPAPPPSP	SAPP	819
Db	1130	KSPPPESVDPTSTKQWPKRSLKADVEEBEFLALRLKLT	P--SACKAMLTPKPA-----	1180
Qy	820	PVPVPEKEKESETAAAPVE-----EGEEQKPPAAEBLA-----	VDTGKA	858
Db	1181	---GGDEKDIKAFMGTPVQKLDLAGLPGSKROLQTPKEKAQALEDLAGKELFQTPGHT	1237	
Qy	859	EEPV-----KSEC---TEEAEGSPAKGKD---AEAAEATAEGALKAEEK---EGSG	GRA	903
Db	1238	EELVAAGKTTIKCDSPQSPDVTPTSTKQRPKRSRKADVEGELLACRNLMP	SAGKAMH	1297
Qy	904	TTAKSGAPOQSDSSATCSADEVDEAEG--GDKNRLLSPRSLTTPGDPRANASQ	PKPL	961
Db	1298	TPKPSVGEKDIIFVGTPTVQKLDLTENLTGSKRR-----	FQTPK	1337
Qy	962	DLQOLKQRAAAIPPIQVTKVHEPRDEAATKPAFPAPPPQNLQPEDAPQPGSS	PRG	1021
Db	1338	EEAQALEDLTGFKEKELFQTPGHTBEAAVAAGKTTKMPCESSPPES-----	ADTPTS	1386
Qy	1022	KSSPAPPADKEAFABE-----AQKLPGDPPCWTSGLPPVPVPPREVIKA	1065	
Db	1387	TRQPKPTLEKRDVQKELSALKKLTQSGTTHTDKVPGGEDKXINAF-----	RETAK-	1439
Qy	1066	SPHAPDPSAFSAPPQHP-----LPLGLHDTA-----RPVLPRPTTISNPPPLI	SSAKH	1114
Db	1440	--OKLDPAASVTSGRKPKTKERKAQPLEDLAGWKELFQTPVCTDKPTTHEKTT	KIACRSQ	1497
Qy	1115	PSVLEROIGAISQMSVOLHVPISE-----HAKAPVGPVTMGLPLP--MDPK	LALPFSGV	1167
Db	1498	PDPVDTPTSSKQSKRSLRKVDVEEBEFALRKETPSAGKAMHTPKPAVSGEKNI	YAFMGT	1557
Qy	1168	KQBOLS--PRQAGQPESLGVPTAQEASVLRGTAIGSVPGSSI	TKGIPSTRVPSDSAITYR	1226
Db	1558	PVKQLDUTENLTGSKRRLQTPKE-----KAQALEDLAG--FKELFQTRGHT	TEE-----	1603
Qy	1227	GSITHGTPADVLYKGTITRIIGBDSPSRLDRGREDLPGKHVITYEGKGHVLS	YEGGMSV	1286
Db	1604	--SMTNDKTA	VACKSQPDL--DKNPASSKRLKTSLGK-----VGKEELLAVG--KL	1652
Qy	1287	TQCKEDGRSSGPPHE-----TAAPKRTYDMGRVGRGAISSASIEGLMGRAI	PPER	1339
Db	1653	TQTSGETTHTBFTGDKSMKAFMESPQILDSAA	SLTG---SKRQLRTPKGSVPED	1709
Qy	1340	-----HSPHLKEQHIRGSIT--QGIPRSYVEAQEDYLREAKLKRECTP	PPPPP	1389
Db	1710	LAFIELFQTPSHKE-----SMTNEKTTKVS	YRASQPD-----LVDTPTSSKQOP	1755
Qy	1390	SRDLTEAYKTQALGPLKPAHEGLVATVKE---AGRSIHEIPREELRHTPEL	LAPRPL	1446
Db	1756	KRSIRKA-----DTEEBFLAFKQTPSACKAMH-----	TPKPA	1788
Qy	1447	-----KSGSITQTPPL--KYDTGASTTGSKKH-----DVRSLIGSPGHT	FFPP	1486
Db	1789	VGEKDINTPLGTPVKQLDQPGNLPGNSRRLQTRKEKAQALEELTG	FRELFQTPCNDNPT	1848
Qy	1487	V-----HPLDVMADA-----RALERACYEESLKS	RCTASSGGS	1526
Db	1849	ADEKTTKKILCKSPQSDPADTPTNTKQRPKRSUKKADVEEFLAFKRLT	PSAGKAMHTPK	1908
Qy	1527	PVTVPE-----LGKPROSLTYEDHAPAGHLPRGSPVTWREPT	PRLQEGSLSSKAS	1580
Db	1909	AAVGEKDINTFVGTPEVK-----LDLLGNLP--GS--KRPQTPK-----	EYAKAL	1951
Qy	1581	QD-----RKLTSPT-----REIA--KSPHSTVPEHHPIPSFYEHLLRGV	SGVDL	1623

Db 1952 EDLAGKELFQTPGHTESMTDDKITEVSKSPQ---EDPVKTPPTSQRKLKISLGKGV 2008
QY 1624 YRSHIPLA-FDPTSIPRGIPLDAAAAYLPRLHAPNPTYPHYLYRGPYDPALENR 1682
Db 2009 KEVLPVGLTQTS-----GKTTQTH-----R 2030
QY 1683 QTIINDYITSQOMHNTATAMQADMLRGL-----SPRESSALNAAAGPRGIIDLSQV 1737
Db 2031 ETAGDG--KSIKAFKESAKQMLDPANYGTGMERWPRTPKEBAQSLEDLAGFK---ELFQT 2085
QY 1738 P-HLPVLPPTGCTATANDRLAYLPTAPQPPSSRHSSPLSPGQPTHLTKPTTSSSER 1796
Db 2086 PDHT-----EESTDDKTKTACKSP-----PEMSDPTST-----R 2118
QY 1797 ERDRERDRDREREKSI---LTSITTTVEHAP-----IWRPCTEISSGSSGSGGG 1845
Db 2119 RPKPTPLGKRDIVEELSALKQITQTHDKVPGDEKGINVRETAKQLDPAASVTG-- 2176
QY 1846 GSSRPASHSHAHQHSPIRPTQDALQORPSVLHNTGMKGII-TAVEPSKPTVLRSTSTS 1904
Db 2177 -----SKQPRTPKGAQPLEDL--AGLKELFQTPVCTDKPTTHEKT-TX 2218
QY 1905 SPVRPAANTPPATHCPGLGTLDCVPTLMEP-----VLLPKEAPRVARP- 1948
Db 2219 IACRSFPQDPVGT-----PTTFKPSQKSLRKADVEESLALRKRTSPVSKAM 2266
QY 1949 ERPRADTG-----HAPLAKPPARSGLSPASPSKSGSEPRPLVP-----PVSCHATIA 1995
Db 2267 DTPKAGGDKMKAFTMGTPVOKLDL-PGNLP--GSKWPQTPKEKAQLEDLAGFKELF 2323
QY 1996 RTPAKN-----LAPHASPPPPAPPASADPHREKTSQSPFSIQE--LELRLSGY 2043
Db 2324 QTPGDKPTTDBKTKTACKSPQDPVDTASTK--QRPKRLKADVEEFLALR---- 2377
QY 2044 HGSSVSPGVEPVPSPVSPSLTHDKGLPKHLEELDKSHLEGELRPKQCPKYLGEAAHL 2103
Db 2378 ---KTPSAGKAMD--TPPAVSDEKNINTFET-----PVOKLDLGLNL 2417
QY 2104 PHRLPSPQSSPSSPLLOTAPGVK-----GHQRVVTLAQHISEVITQD-----YTRH 2150
Db 2418 PGSKRQPTPKREKAEALDVLGFKELFQTPGHTESMTDDKITEVSKSPQSPESFKTSR 2477
QY 2151 HQQLSAPLAPLYSPGASCPVLDRPP---SDLYLPPPD---HGAPARGSPHSEGG 2203
Db 2478 SKQRLKPLVK-----VDMKEPLAVSKLRTSGTQTTHTPTGDSKSIKAF 2525
QY 2204 KRSPB---PNKTSVLGGGED-----GTEP-VSPPEGMTEPCHRSRVAVPLLYRDE 2250
Db 2526 KESPKQILDPAASVTGSRQRLTRKEKARALEDLVDFKELFSAPGHTESM-----T 2577
QY 2251 QTEPSRMGSKPGNTSQPPAFPSKLTESNMAVSKKOEINKKLNTHRNPEYNI SQPG 2310
Db 2578 IDKNTKPKCKSP-----PPELTDATSTKRCPTKPRKEVKEELSAVERLT---QTSGQS 2629
QY 2311 TEIFNMPAITGTGLMYRQA-----VOEHAS----- 2337
Db 2630 THTHKEPASGDEGKVLKQRAKKPNPVEEPSPRRRPPAPKEKAQPLEDLAGFTLSETS 2689
QY 2338 --TNMGLEAIRKALMGKYDOWE--ESPPLSANAFNPLNASAL-----P 2378
Db 2690 GHTQESLTA-----GKATKIPESPPELV-----VDTTASTYRHLTRVQKVQVKEEP 2737
QY 2379 AAMPITAADGRSDHMLTSPGGGKAKVSGRPSRRKAKSPAPGLASGRDP 2428
Db 2738 SAVKFTQTSGETTDADKEPAGEDKGIKALKESAKQTAPAPASVTGSRRRP 2787

RESULT 59

ID ACIN_MOUSE STANDARD; PRT; 1338 AA.
AC Q9JIX8; Q9CSN7; Q9CSR9; Q9CSX7; Q9R046; Q9R047;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)

10-OCT-2003 (Rel. 42, Last annotation update)
Apoptotic chromatin condensation inducer in the nucleus (Acinus).
ACINUS.
Mus musculus (Mouse).
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI TaxID=10090;
[1]
SEQUENCE FROM N.A. (ISOFORMS 2 AND 3).
MEDLINE=99418556; PubMed=10490026;
Sahara S., Aoto M., Eguchi Y., Imamoto N., Yoneda Y., Tsujimoto Y.;
"Acinus is a caspase-3-activated protein required for apoptotic
chromatin condensation.";
Nature 401:168-173(1999).
[2]
SEQUENCE FROM N.A. (ISOFORM 1).
Manoru A., Setsuko S., Yoshihide T.;
"Molecular cloning of murine acinusl, a gene for apoptotic chromatin
condensation.";
Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 1-1190 FROM N.A. (ISOFORMS 1; 2 AND 4).
STRAIN=CS7BL/6J; TISSUE=Embryo, and Pancreas;
MEDLINE=22354683; PubMed=12466851;
Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
Willing L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
Birney E., Hayashizaki Y.;
"Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
Nature 420:563-573(2002).
CC -!- FUNCTION: Induces apoptotic chromatin condensation after
activation by CASP3 (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=4;
CC Name=1; Synonyms=1;
CC IsoId=Q9JIX8-1; Sequence=Displayed;
CC Name=2; Synonyms=8;
CC IsoId=Q9JIX8-2; Sequence=VSP_004030, VSP_004033;
CC Name=3; Synonyms=S';
CC IsoId=Q9JIX8-3; Sequence=VSP_004031;
CC Name=4;
CC IsoId=Q9JIX8-4; Sequence=VSP_004032;
CC -!- PTM: Undergoes proteolytic cleavage; the processed form is active,
contrary to the uncleaved form (By similarity).
CC -!- SIMILARITY: Contains 1 SAP domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to frameshifts
in position 110 and 112.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the ENBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sb-sib.ch).

ENBL; AFI24725;	AAD56723.1; -
ENBL; AFI24729;	AAD56727.1; -
ENBL; AFI68782;	AAF89661.1; ALT FRAME.
ENBL; AK011698;	-; NOT ANNOTATED_CDS.
ENBL; AK012099;	BAB28030.1; -
ENBL; AK012337;	BAB28171.2; -
ENBL; AK050467;	BAC34272.1; ALT_INIT.
MCD; MGI:1891824;	Acinus.
GO; GO:0005730;	C:nucleolus; IDA.
GO; GO:0016887;	F:ATPase activity; ISS.
GO; GO:0019899;	F:enzyme binding; ISS.
GO; GO:0003676;	F:nucleic acid binding; ISS.
GO; GO:0003263;	P:apoptotic chromosome condensation; ISS.
GO; GO:0030218;	P:erythrocyte differentiation; ISS.
GO; GO:0045657;	P:positive regulation of monocyte differentiation. . ; ISS.
InterPro; IPR003034;	SAP.
Pfam; PF02037;	SAP; 1.
SMART; SMO0513;	SAP; 1.
PROSITE; PS00800;	SAP; 1.
KW Apoptosis;	Nuclear protein; Alternative splicing.
FT DOMAIN	72 106 SAP.
FT DOMAIN	142 443 GLU-RICH.
FT DOMAIN	569 667 SER-RICH.
FT DOMAIN	1113 1130 PRO-RICH.
FT DOMAIN	1131 1338 ARG/ASP/GUU/LYS-RICH.
FT SITE	1093 1094 CLAVAGE (BY CASPASS-3) (BY SIMILARITY).
FT VARSPLIC	1 757 Missing (in isoform 2).
FT VARSPLIC	1 773 /FTId=VSP_004030.
FT VARSPLIC	164 204 Missing (in isoform 3).
FT VARSPLIC	758 766 /FTId=VSP_004031.
FT VARSPLIC	244 244 EASAESDSTHPEGVASLLPPDFQSSLNRPPELSTHSPR
FT VARSPLIC	515 515 -> G (in isoform 4).
FT VARSPLIC	536 536 /FTId=VSP_004032.
FT VARSPLIC	557 557 ESERTHTVT-> MMFSDSRAG (in isoform 2).
FT VARSPLIC	568 568 T-> A (IN REF. 2).
FT VARSPLIC	569 599 T-> A (IN REF. 2).
FT VARSPLIC	599 599 F-> L (IN REF. 2).
FT VARSPLIC	729 729 G-> D (IN REF. 2).
FT VARSPLIC	757 757 H-> Y (IN REF. 2).
FT VARSPLIC	773 773 V-> A (IN REF. 2).
FT VARSPLIC	829 829 S-> I (IN REF. 2).
FT VARSPLIC	896 896 L-> P (IN REF. 2).
FT VARSPLIC	896 896 T-> A (IN REF. 1; AAD56723).
FT VARSPLIC	1035 1035 MISSING (IN REF. 3; BAB28030).
FT VARSPLIC	1338 AA; 150690 MW; B912DSCB5750BPB4 CRC64; Q-> R (IN REF. 3; BAB28030).
FT VARSPLIC	G-> R (IN REF. 3; BAB28030).
FT VARSPLIC	SEQUENCE 1338 AA; 150690 MW; B912DSCB5750BPB4 CRC64;
Query Match	2.6%; Score 346; DB 1; Length 1338;
Best Local Similarity	19.1%; Pred.No. 0.00016;
Matches 329;	Conservative 206; Mismatches 570; Indels 614; Gaps 74;
QY	250 POVELPLYNOPSDTROVHENIKIN-----QAMR-KKILYFKRRHARKQKKFCORY 302
DB	43 PKMPEAVGTDPSRSRWAALEEVTLDGKPLQALRVDTLKAALSGRLAKSGQSALVKRL 102
QY	303 D--QLMEALEK-KVERIENNPRRAKESKVREYYEKQFPPIRKORLQERMQRSGVRGS 359
DB	103 KGALMLLENLKQSTHPAAFPNSQIGEMSONSFQKY--LEKQELLRLRLEREAREAA 160
QY	360 GLSWSAARSEHYSEIIDGLSE-----QENLEKQMRLQAVIPMLPYDADQQRIKFNNW 413
DB	161 ELBEASAESDEWTH-PEGVASLLPPDFQSSLNRPPELSTHSP----- 203
QY	414 GLMADPMVKYDRQVMNMMSQEKEPTPREKFMQKNFGLIASFLEKTKVAECVLYYLT 473

```
Db 987 GVSITIDDPVTAQVSPSPKIGKSNIVHISNLVRPFTLLQKELLGRTGTLVEAFWIDK 1046
QY 1459 -----YDT--GASTTCKSKKHVRSILGSPRTFPVPHPLDVWADARALACIYEESL 1508
Db 1047 IKSHCFVTYSTVEEAVATRTALHGVK-----WQSNPKFLCADYAEQDELHYH--- 1094
QY 1509 KSRPGTASSGGSIARGAPVIVPELKGKQPSPLTYEDHGAPFAGHLPRGSPVTWREPTP- 1567
Db 1095 -----RGLLV-----DRPSETKAEQAGAPRPLHPPPPPV---QPPPH 1129
QY 1568 -----RLQSGUSSKASQDR-----KLSTPR-----ETAK 1594
Db 1130 PRAEQEGERAVREQWAEREREMERRTSRERWDRKVREGPRSRKSRDRRKRERAK 1189
QY 1595 SPHSTVPEHHPHPISPYEHLLRGVSGVLYRSHIPLADPTSPICGIPLDAAAYLPRH 1654
Db 1190 SKEKSEKKEKAQEAPPAKLID-----DLFRK-----TKAAPCIYWLDP-- 1227
QY 1655 LAPNPTYPHLYPIRLGYPDTAALENQTIINDYITSQQMHNTATAMAQRADMLRGLS 1714
Db 1228 -----LTSQIVQKEAQERAKERKRRK 1252
QY 1715 PRESSLALNAAAGPGRIIDLSQVPHLPVLVPTPGTATAMDRLAYLPTAPQPFSSRHSS 1774
Db 1253 EREEE-----EQER-----EKEAERERN 1271
QY 1775 SPLSPGGTHLTKPTTSSSRERDRDRDR-DREREK 1812
Db 1272 RQLE-----REKRSRERDRERDRGRDRER 1303

RESULT 60
APC_MOUSE
ID APC_MOUSE STANDARD; PRT; 2845 AA.
AC Q61315; Q62044;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Adenomatous polyposis coli protein (APC protein) (mAPC).
GN APC.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS.
RP STRAIN=CS7BL/6J, and CAST/EI; TISSUE=Brain;
RX MEDLINE=92263101; PubMed=1350108;
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RT "Multiple intestinal neoplasia caused by a mutation in the murine
RT homolog of the APC gene.";
RL Science 256:668-670(1992).
RN [2]
RP ERRATUM.
RA Su L.-K., Kinzler K.W., Vogelstein B., Preisinger A.C., Moser A.R.,
RA Luongo C., Gould K.A., Dove W.F.;
RL Science 256:1114-1114(1992).
RN [3]
RP SEQUENCE OF 1-45 FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RA Dicker F., Lambert S., Reitmaier A., Ballhausen W.G.;
RT "The murine APC gene: alternative splicing of 5' untranslated
RT region segments.";
RL Submitted (OCT-1993) to the EMBL/GenBank/DBJ databases.
RN [4]
ALTERNATIVE SPLICING.
RX MEDLINE=94061824; PubMed=8242607;
RA Oshima M., Sugiyama H., Kitagawa K., Taketo M.;
RT "APC gene messenger RNA: novel isoforms that lack exon 7.";
RL Cancer Res. 53:5589-5591(1993).
CC -!- FUNCTION: Tumor suppressor. Promotes rapid degradation of CTNNB1
and participates in Wnt signaling. APC activity is correlated with
```

```
CC CC its phosphorylation state (By similarity).
CC CC -!- SUBUNIT: Forms homooligomers. Associates with catenins. Binds
CC CC axin (By similarity).
CC CC -!- ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=4;
CC CC Name=1;
CC CC IsoId=Q61315-1; Sequence=Displayed;
CC CC Name=2; IsoId=Q61315-2; Sequence=VSP_004116;
CC CC Name=3; IsoId=Q61315-3; Sequence=VSP_004117;
CC CC Name=4; IsoId=Q61315-4; Sequence=VSP_004116, VSP_004117;
CC CC -!- TISSUE SPECIFICITY: Expressed in liver, spleen, kidney, heart,
CC CC lung, brain, stomach, intestine, testis and ovary.
CC CC -!- PTM: Phosphorylated by GSK3B (By similarity).
CC CC -!- SIMILARITY: Contains 7 ARM repeats.
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; M88127; AAB59632.1; -.
CC CC EMBL; U02937; AAA03443.1; -.
CC CC PIR; I49505; I49505.
CC CC HGSP; Q02248; 3BCT.
CC CC MGD; MGI:880339; APC.
CC CC GO; GO:0005737; Cytoplasm; IDA.
CC CC GO; GO:0005634; C:nucleus; IDA.
CC CC GO; GO:0008013; P:beta-catenin binding; IDA.
CC CC GO; GO:0009952; P:anterior/posterior pattern formation; IMP.
CC CC GO; GO:0009798; P:axis specification; IMP.
CC CC GO; GO:0009953; P:dorsal/ventral pattern formation; IMP.
CC CC GO; GO:0016055; P:Wnt receptor signaling pathway; IDA.
CC CC InterPro; IPR008938; ARM.
CC CC InterPro; IPR000225; Armadillo.
CC CC Pfam; PF00514; Armadillo_seg; 4.
CC CC SMART; SM00185; ARM; 5.
CC CC PROSITE; PS50176; ARM_REPEAT; 1.
CC CC Wnt signaling pathway, Anti-oncogene; Phosphorylation;
CC CC Alternative splicing; Repeat; Coiled coil (POTENTIAL).
CC CC DOMAIN 1 61 COILED COIL (POTENTIAL).
CC CC DOMAIN 125 245 COILED COIL (POTENTIAL).
CC CC REPEAT 1 728 LEU-RICH.
CC CC REPEAT 451 493 ARM 1.
CC CC REPEAT 503 545 ARM 2.
CC CC REPEAT 546 589 ARM 3.
CC CC REPEAT 590 636 ARM 4.
CC CC REPEAT 637 681 ARM 5.
CC CC REPEAT 682 723 ARM 6.
CC CC REPEAT 724 765 ARM 7.
CC CC DOMAIN 739 2834 SER-RICH.
CC CC DOMAIN 1130 1156 ASP/GLU-RICH (ACIDIC).
CC CC DOMAIN 1556 1575 ASP/GLU-RICH (ACIDIC).
CC CC DOMAIN 1864 1891 HIGHLY CHARGED.
CC CC VARSPLIC 243 276 Missing (in isoform 2 and isoform 4).
CC CC VARSPLIC 310 410 Missing (in isoform 3 and isoform 4).
CC CC VARSPLIC 120 410 Missing (in isoform 3 and isoform 4).
CC CC VARIANT 120 410 T -> A (IN STRAIN CAST/EI).
CC CC VARIANT 493 493 Y -> I (IN STRAIN CAST/EI).
CC CC VARIANT 797 797 Y -> F (IN STRAIN CAST/EI).
CC CC VARIANT 1330 1330 A -> T (IN STRAIN CAST/EI).
CC CC VARIANT 1618 1618 A -> S (IN STRAIN CAST/EI).
CC CC VARIANT 2294 2294 G -> A (IN STRAIN CAST/EI).
CC CC VARIANT 2496 2496 H -> Q (IN STRAIN CAST/EI).
CC CC VARIANT 2523 2523 T -> A (IN STRAIN CAST/EI).
CC CC VARIANT 2813 2813 T -> S (IN STRAIN CAST/EI).
CC CC SEQUENCE 2845 AA; 311086 MW; 145CA73CF570A99 CRC64;
```



```
Db 243 GOIQSGAAQMAQETRDALKCDVTSALRE---IRAQLEGHAVOSTLQSEWFRVRLDR 299
Qy 305 LMEALEKKVRIENPRRRRAKESKVEYYEQFPEIRKQRELQBMQSRVQQRGSLMS 364
Db 300 LSEAAK-----VNTDAMSQAQETIY-----RRQLQARTTELEALKSTKOSLE 343
Qy 365 AARSEHEVSEIIDGLSEQENLEKOMROLAVIPMLYDADQOQRIKFINNGIAMDPMKYK 424
Db 344 RQSELEDRHODIASYQEAITQ-----LQAELRNTKW-----EMAAQLREYQ 386
Qy 425 DRQVMNMWSEOEKETFREKFWHPK---NFGLIASFLEKRTVAECVLVYYLTKNENYKS 481
Db 387 DLLNVKMWALDIEIAAYR-KLLEGEBCRIGFGPIFSLPEGLPKIPSPVSTHIKXSEKIK 445
Qy 482 LVRSYRRRGKSQQQQQQQQQQQQQQPPRSQEQEKDEKEKEAEKEEKEEPEVEND 541
Db 446 VVESEKETVIVEQTEETQTEEV-----TEEEKEAKEBEGKEEGEGBEEAEGG 497
Qy 542 KEDLLKEKTDGTDGDNDEKEAVASKGRKTKANSQGRKGRITRSMANEANSEEAITP--- 598
Db 498 BEETKSPPAEBAAGPEAKSPVKEAKSPAEAKSPEKEE-AKSPAEVKSPKAKSPAKE 556
Qy 599 -----QOSAEIASMELNNESSRWTE-----EMETAKKGLLEHGRNWS 635
Db 557 BAKSPPEAKSPEKEAKSPAIVKSPKAEKSPAKEAKSPAEAKSPEKAKSPVKEAKSPA 616
Qy 636 AIARMVGSKTQSQCKNFYFNKKQNLDEILQOHLKM-----EKERNARRKKKAPA 688
Db 617 EAKSPVKEAKSPAIVKSPKAEKSPTEKEAKSPEKAKSPEKAKSPEKEAKSPV 676
Qy 689 ---AASBEAAPPPVVEDEMBASGVSGNEEEMVEAEALHASGNEVPRGECSPATVNN 745
Db 677 KAEAKSPEKAKSPVKAEBKSPKAEKSPVKEAKSPEKAKSPEKAKSPVKEA 736
Qy 746 SDTESISPTEAAKQDQ-KNPPRPAATLGDGPPGPPPTPRTSRAPTEPTASEATG 804
Db 737 KTEPAKSPVKEAKSPEKAKSPEKAKTL--DVKSPEAKTPAKBEAKSPADKFP--EKAK 792
Qy 805 APTPPAPPPSPAPPPVVPKEKEEETAAAPVVEEGEBQKPPAAELAVDTGKAEPEVKS 864
Db 793 SPVKEEVKSPKAKSPL-----KEDAKAPEKEIKPEKEVKSPVKEEKPOEVKEPPKK 847
Qy 865 ECTEAEAGPAKGDAAEAATAEGALKAKEKGGSGRATTAKSGAPQDSDSATCSAD 924
Db 848 ---AEEKAPATPKTEKKDSKE---EAPKKEAPKPKVEKSGPAVEKPKSKV----- 896
Qy 925 EVDEAEGDKRLLSPRSLTPTGDPANASPKPLDLKQKORAAAIPPIQVTKWHEP 984
Db 897 EAKKEAEDKKV-----PTPEKEAPAKVEKEDAK-----PKETEVAKK 937
Qy 985 PRED---APTKAPPAPPPQNLQPSDAPQPGSSPRGK-----SRSPA-PPA 1030
Db 938 EPDDAKAPEKSPKAEKPAEAKPEKDTKEKAKPKTEKAKEDKTLKSPKPKPA 997
Qy 1031 DKEAFAPAEAKLPGDPP 1047
Db 998 EAKESSTQDKSKPP 1014

RESULT 62
Call MOUSE
ID Call MOUSE STANDARD; PRT; 1453 AA.
AC P11087; O60635;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
GN COL1A1 OR COLA1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
```

```
RP RC SEQUENCE FROM N.A.
RX STRAIN=FVB/N;
RX MEDLINE=96033240; PubMed=8535610;
RA Li S.W., Khillan J., Prockop D.J.;
RT "The complete cDNA coding sequence for the mouse pro alpha 1(I) chain
of type I procollagen.";
RL Matrix Biol. 14:593-595 (1995).
RN [2]
RP RC SEQUENCE OF 518-1128 FROM N.A.
RX MEDLINE=86137403; PubMed=3841523;
RA French B.T., Lee W.-H., Maul G.G.;
RT "Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)
collagen protein.";
RL Gene 39:311-312 (1985).
RN [3]
RP RC SEQUENCE OF 735-1130 FROM N.A.
RX MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1 (I) procollagen gene:
evidence for a mouse B1 element within the gene.";
RL Mol. Cell. Biol. 2:1362-1371 (1982).
RN [4]
RP RC SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
RX MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a Balb/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69 (1981).
RN [5]
RP RC SEQUENCE OF 1442-1453 FROM N.A.
RX MEDLINE=88124276; PubMed=3340560;
RA Mooslechner K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
of the 3'-untranslated region.";
RL Nucleic Acids Res. 16:773-773 (1988).
CC -!- FUNCTION: type I collagen is a member of group I collagen
(fibrillar forming collagen).
CC -!- SUBUNIT: Trimers of one alpha 2(I) and two alpha 1(I) chains.
CC -!- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
bones. In bones the fibrils are mineralized with calcium
hydroxyapatite.
CC -!- PRT: Prolines at the third position of the tripeptide repeating
unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- SIMILARITY: Contains 1 VWF domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U08020; AAA89912.1; -
DR EMBL; X15896; CAA33904.1; -
DR EMBL; M14423; AAA37333.1; -
DR EMBL; M17491; AAA37334.1; -
DR EMBL; X06753; CAA29927.1; -
DR EMBL; K03036; AAA37332.1; -
DR EMBL; K03029; AAA37332.1; JOINED.
DR EMBL; K03030; AAA37332.1; JOINED.
DR EMBL; K03031; AAA37332.1; JOINED.
DR EMBL; K03032; AAA37332.1; JOINED.
DR EMBL; K03033; AAA37332.1; JOINED.
DR EMBL; K03034; AAA37332.1; JOINED.
DR EMBL; K03035; AAA37332.1; JOINED.
DR PIR; S57243; S21626.
DR MGI; MGI:88467; Coll1a1.
DR InterPro; IPR008161; C1q helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib-collagen_C.
DR InterPro; IPR002181; Fibrinogen_C.
DR InterPro; IPR001007; VWF_C.
```

DR Pfam; PF01410; COLFI; 1.
 DR Pfam; PF01391; Collagen; 18.
 DR ProDom; PD000007; C1g helix; 1.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VMC; 1.
 DR PROSITE; PS01208; WFCC 1; 1.
 DR PROSITE; PS0184; WFCC 2; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 22
 FT PROPEP 23 151
 FT CHAIN 152 1207
 FT PROPEP 1208 1453
 FT DOMAIN 29 87
 FT DOMAIN 152 167
 FT DOMAIN 168 1181
 FT DOMAIN 1182 1207
 FT CARBOHYD 56 56
 FT CARBOHYD 1354 1354
 FT SITE 734 736
 FT SITE 1082 1084
 FT CONFLICT 1450 1450
 FT SEQUENCE 1453 AA; 137944 MW; 38802535DF81808 CRC64;
 Query Match 2.6%; Score 341; DB 1; Length 1453;
 Best Local Similarity 21.7%; Pred. No. 0.00025;
 Matches 335; Conservative 89; Mismatches 539; Indels 580; Gaps 80;
 QY 707 ASGVSGNHEEMVEAEALHASQNEVPRGE-----C-SGPAVNNSSDTSIESPSP 754
 DB 15 ATALLTHQOEIDPEVSCIH-NGLRVFNGETWKPEVCLICICNGHTAVCDVQCNEELDCP 73
 QY 755 HTEA-----AKDTGQNGPK-----PPTLGDGPP--PGP 782
 DB 74 NPQRREGCCACFPREYVSPNSDVGVEGKGGPQPGPQPGVPPGRDGIIPGQGLPGP 133
 QY 783 PTPPRTSRAPTEPTPASEATGATPPP-----PAPSPSPAP- 818
 DB 134 PGPP-----GPPPGPLGGNFASQMSYGVYDEKSAQSVFPGMGPSPGR 176
 QY 819 ---PPVVKKEKEBETAAPVEGE-----EQKPPAAELAVDTGKAEPVVKSEC 866
 DB 177 GLPGPPGAPGQFGQ-----GPPGEPGPGGMPGPPGPPGKNGDGDGAGKPGR--- 229
 QY 867 TEEAEGPAKGDAAEATAEGALKAKEKGGSGRATTAKSSGAPQDSDSATCSADEV 926
 DB 230 ---PGERGPPQAGRLPCTA--GLPGMKHGRGSLDGAKGDAGP----- 271
 QY 927 DEAEQDKNRLSP-----RPSLLTPTGDPANASPKPLDLKQLKQRAAAIPPIQVTKVH 982
 DB 272 ---AGPKGEPGSPGNGAPGQMGPRGLPGERGRGPP-----GTAGA-----RGN 313
 QY 983 EPPREDAAPTKAPAPP-----PQNLPQESDAPQ-----QPSSPRGKRS 1025
 DB 314 DAGVAAGPGPTGTGTPGPFPGFVAGKAGAGPQAR--GSEGPQGVGEPGPPGAGAA 372
 QY 1026 PA--PPADKEAFAEAQKLPDPPCWTISGLP-FVPPPREVIKASPHAPDPSAFYAPPGH 1082
 DB 373 PAGNPGADGQPGAKGANGAPG-----IAGAPGFCARGPSPGQPSGPPGKNGSGPFGA 427
 QY 1083 PLPLGLHDTARPVLPPTPTINNP-PLISSAKH-----PSVL-----ERQIGAISQGM 1129
 DB 428 PGNKG--DITGAKGFGATGVGQPGAPAGEGKRGARGEPGPGSLPFGPPGERG--GPGSRGF 484
 QY 1130 SVQLHVPYSEHAKAPVGVV--TWGLPLPMDPKLAPFSVQKEQLSPRGQAGPPESLVPT 1188
 DB 485 -----FCADGVAGKPGSGERGAPGAPGPKG-----SP-GEAGRPGCEAGLPG 525
 QY 1189 AQEASVLRGT-----ALGSVPGGSIKGIPTSPVPSDAITVRGSI--THGTADVLYKGTI 1243
 DB 526 AKGLTSGSPGPGDKTGTGPPGAGQDGRGPPGAPPGA-----RGQAGVMGFFGP---KGT- 577

QY 1244 TRIIGEDSPSL-DRGREDSLKPHGVIEGKGKGVLSVEGMSVTQCSKEDGRSSSGPPH 1302
 DB 578 ---AGE--PGKAGERGLPG--PPGAVGPAGKDG-----EAGAQAAGPAGPAG 618
 QY 1303 ETAAPKRTYDMMEGRVGRAISSASIEGLMGRAIIPP-ERHSPHILKEQHHRGSIQTQIPR 1361
 DB 619 E-----RGEQGA--GSPGFQGLPGPAGPPGEAGKPE-----QGVPG 654
 QY 1362 SYVEAQEDYLRRKLLKREGTPPPPPPSRDLTEAYKTQALGPL--KLKPAHEGLVATVK 1419
 DB 655 DLGAPGPGSGARGERGPPGQVQPPGPA-----GPRNGNAGPNDCAKGDGTG 702
 QY 1420 EAGRSIHEIPREELRHTELPPLAPRLKEGSIQTGP-----LKDYDTGASITGSKHP--- 1472
 DB 703 APG-----AP-----GS--QGAFLQGMPPGERGAAGLPGPKGDRGD 736
 QY 1473 --VRLSLGSPGR-----TFP--PVHPLDMADARALERACYEBSLSRRPQTASSGCSI 1522
 DB 737 AGPKGADGSPKGCGARGLTGPIGPPGAPAGD-----KGEAGPSGPPGPTG 783
 QY 1523 ARGAPVIVPELGRQPSRLTYEDHGAPFAGHLPRGSPVTWREPTPRLOEGLSSSKASQD 1582
 DB 784 ARGAPGDRGEAGPP--GPAGFA--GPPGAD-----GQFKAKEGPDGTGVKGD----- 826
 QY 1583 RKLTSPTREIAKSPHSTVPEHHPHPI---SPVEHLLRGVGVLDLYESHIPLAFDPTSIP 1638
 DB 827 -----AGPPGAPGAPGPGPIGNVCAAPGPKGRGAAP-----P 860
 QY 1639 RGIPLDAAAAYILPRHLPNPTYPHYLYPIRIGYDPTAALENROTIINDYITSQQMHN 1698
 DB 861 GATGFPGAAGRVG---PGFS--GNAGPP---GPPGPVCKEG----- 894
 QY 1699 TATAMAQADMLGSLSPRESSLALNVAAGPRGIIIDISQVPHLPVLVPTPTGTATAMDRL 1758
 DB 895 -----GKGRGET-----GPAG-----RPEVGP-EGPPFPAGEKG 924
 QY 1759 AYLTPAPQPPSSRHSSSPLSPG-----GPTHLTKPTTTS 1792
 DB 925 S--PGADGP-----AGSPGTPGQGIAGQGVVGLPQRCGERGFPGLPGSPGPKQGPS 977
 QY 1793 SSERDRDRDRDREREKSIILTSITTVHAPIWRPGTEQSSGSSGSSG-----GGGGS 1847
 DB 978 GSSGERG-----PPGMPGPPGLAGPPGCSGREGSPGAEKSPGR 1015
 QY 1848 SSRPASHSHAHQSPISPTQDALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSSPV 1907
 DB 1016 DGAPGAKGDRGETGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPGAPG 1067
 QY 1908 RPAATFPPTATHCPGLGTLGVYPTLMPEVLLPKPEAPRVARPERPRADTGH----- 1957
 DB 1068 GPAGARGP-----AGPQGRGDKGETGQDGRGIK 1097
 QY 1958 -----AFLAKPPARSL-----EPASSPSKGE-----PRPLVPFVSG 1990
 DB 1098 GHRGFSGLQPPGSPGSGQSGASGAPGPRGPPGSGAGSPGKGLNGLPGIPGPPGR 1157
 QY 1991 HATIAITPAKNLAPHASDDPPAPPASAD-----PHREKTO 2027
 DB 1158 GRTGDSGPAGPPCP-PGPPGPPGPPSGGVDFFSLPPOPPEKSQ 1199
 RESULT 63
 SHK3 RAT
 ID SHK3 RAT STANDARD; PRT; 1815 AA.
 AC OSULIA; Q9WU7; Q9WU47;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE SH3 and multiple ankyrin repeat domains 3 (Shank3) (Proline-rich
 DE synapse associated protein 2) (PROSAP2) (SPANK-2).
 GN SHANK3.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

[1] NCBI_TaxID=10116;

SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH DLGAP1 AND DLG4.

MEDLINE=9458653; PubMed=10527873;

RA Boeckers T.M., Winter C., Smalla K.-H., Kreutz M.R., Bockmann J.,

RA Seidenbecher C., Garner C.C., Gundelfinger E.D.;

RT "Proline rich synapse associated proteins Prosap1 and Prosap2 interact

RT with synaptic proteins of the SAPAP/GKAP family.";

RL Biochem. Biophys. Res. Commun. 264:2476-2528(1999).

[2]

SEQUENCE FROM N.A. (ISOFORM 2), AND INTERACTION WITH DLGAP1 AND CCTN.

RC STRAIN=Sprague-Dawley; Tissue=Hippocampus;

RA MEDLINE=99360650; PubMed=10433268;

RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,

RA Weinberg R.J., Worley P.F., Sheng M.;

RT "Shank, a novel family of postsynaptic density proteins that binds to

RT the NMDA receptor/PSD-95/GKAP complex and cortactin.";

RL Neuron 23:569-582(1999).

[3]

PARTIAL SEQUENCE FROM N.A. (ISOFORM 3).

RC Tissue=Brain;

RA MEDLINE=20549637; PubMed=10958799;

RA Tobaben S., Suedhof T.C., Stahl B.;

RT "The G protein-coupled receptor C11 interacts directly with proteins

RT of the Shank family.";

RL J. Biol. Chem. 275:36204-36210(2000).

[4]

INTERACTION WITH HOMER PROTEINS; DLGAP1; MGLURIA AND MGLUR5, AND

RP MUTAGENESIS LEU-1386 AND CYS-1389.

RA MEDLINE=99360651; PubMed=10433269;

RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,

RA Dean A., Aakalu V.K., Lanahan A.A., Sheng M., Worley P.F.;

RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of

RT postsynaptic density proteins.";

RL Neuron 23:583-592(1999).

[5]

REVIEW.

RA MEDLINE=20267867; PubMed=10806096;

RA Sheng M., Kim E.;

RT "The Shank family of scaffold proteins.";

RL J. Cell Sci. 113:1851-1856(2000).

CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic

CC density (PSD) of excitatory synapses that interconnects receptors

CC of the postsynaptic membrane including NMDA-type and metabotropic

CC glutamate receptors via complexes with GKAP/PSD-95 and Homer,

CC respectively, and the actin-based cytoskeleton. May play a role in

CC the structural and functional organization of the dendritic spine

CC and synaptic junction.

CC -!- SUBUNIT: May homomultimerize via its SAM domain (By similarity).

CC Interacts with DLGAP1/GKAP, MGLUR1A and MGLUR5 C-termini via its

CC PDZ domain. Interacts with Homer-1, Homer-2, Homer-3 and

CC CCTN/cortactin SH3 domain. Is part of a complex with DLG4/PSD-95

CC and DLGAP1/GKAP.

CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of

CC neuronal cells. Extends into the region subjacent to the PSD.

CC -!- ALTERNATIVE PRODUCTS:

CC Name=2; Synonyms=A;

CC IsoId=Q9JLU4-1; Sequence=Displayed;

CC Name=1;

CC IsoId=Q9JLU4-2; Sequence=VSP_006089;

CC Name=3;

CC IsoId=Q9JLU4-3; Sequence=VSP_006087, VSP_006088;

CC -!- TISSUE SPECIFICITY: Predominantly expressed in brain.

CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.

CC -!- SIMILARITY: Contains 4 ANK repeats.

CC -!- SIMILARITY: Contains 1 PDZ/DRH domain.

CC -!- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.

CC -!- SIMILARITY: Contains 1 SH3 domain.

CC -!- CAUTION: Ref.2 sequence differs from that shown in positions 3 to

CC 37 due to a frameshift.

CC -!- CAUTION: Ref.3 sequence differs from that shown in positions 973

CC CC to 1000 due to a frameshift.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>

CC or send an email to license@sib-sib.ch).

CC -----

DR EMBL; AJ133120; CAB45688.1;

DR EMBL; AF133301; AAR61375.1; ALT_FRAME.

DR EMBL; AF159047; AAD42976.1; ALT_FRAME.

DR InterPro; IPR002110; ANK.

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR001660; SAM.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00023; ank; 5.

DR Pfam; PF00595; PDZ; 1.

DR Pfam; PF00536; SAM; 1.

DR Pfam; PF00018; SH3; 1.

DR ProDom; PD000066; SH3; 1.

DR SMART; SM00248; ANK; 5.

DR SMART; SM00228; PDZ; 1.

DR SMART; SM00454; SAM; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS50088; ANK_REPEAT; 4.

DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.

DR PROSITE; PS50106; PDZ; 1.

DR PROSITE; PS50105; SAM_DOMAIN; 1.

DR PROSITE; PS50002; SH3; 1.

DR ANK repeat; SH3 domain; SH3-binding; Coiled coil; Repeat;

FW Alternative splicing.

FT REPEAT 223 256 ANK 1.

FT REPEAT 257 289 ANK 2.

FT REPEAT 324 356 ANK 3.

FT REPEAT 357 389 ANK 4.

FT DOMAIN 472 488 COILED COIL (POTENTIAL).

FT DOMAIN 545 604 SH3.

FT DOMAIN 645 739 PDZ.

FT SITE 1486 1492 SH3-BINDING (POTENTIAL).

FT DOMAIN 1570 1590 COILED COIL (POTENTIAL).

FT DOMAIN 1752 1815 SAM.

FT DOMAIN 6 15 POLY-ALA.

FT DOMAIN 895 901 POLY-PRO.

FT DOMAIN 914 919 POLY-PRO.

FT VARSPLIC 1204 1211 SPTVHSP -> PRRRAGMV (in isoform 3).

FT VARSPLIC 1212 1815 Missing (in isoform 3).

FT VARSPLIC 1612 1620 Missing (in isoform 1).

FT VARSPLIC 1386 1386 Missing (in isoform 1).

FT MUTAGEN 1386 1386 P->L: ABOLISHES INTERACTION WITH HOMER-1

FT MUTAGEN 1389 1389 ISOFORM C.

FT CONFLICT 1 2 MQ -> AA (IN REF. 2).

FT CONFLICT 32 32 P -> Y (IN REF. 2).

FT CONFLICT 37 37 R -> Q (IN REF. 2).

FT CONFLICT 68 68 Q -> P (IN REF. 2).

FT CONFLICT 472 472 H -> L (IN REF. 3).

FT CONFLICT 781 781 R -> G (IN REF. 3).

FT CONFLICT 902 907 YFDSG -> ILRLR (IN REF. 3).

FT CONFLICT 912 917 FSPPPP -> SHGHQ (IN REF. 3).

FT CONFLICT 921 921 R -> G (IN REF. 3).

FT CONFLICT 963 963 R -> G (IN REF. 3).

FT CONFLICT 1115 1115 S -> N (IN REF. 2).

FT CONFLICT 1162 1162 S -> N (IN REF. 2).

FT CONFLICT 1167 1167 G -> S (IN REF. 2).

FT CONFLICT 1337 1337 S -> N (IN REF. 2).

FT CONFLICT 1345 1345 G -> S (IN REF. 2).

FT CONFLICT 1348 1348 S -> N (IN REF. 2).

FT CONFLICT 1354 1354 E -> K (IN REF. 2).

RESULT 64
 ID CAL3 HUMAN STANDARD; PRT; 1466 AA.
 AC P02451; Q15112;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-JAN-1990 (Rel. 13, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Collagen alpha 1(III) chain precursor.
 GN COL3A1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Skin fibroblast;
 RX MEDLINE=89350838; PubMed=2764886;
 RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
 RA Prockop D.J.;
 RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
 RT chain of human type III procollagen. Differences in protein structure
 RT from type I procollagen and conservation of codon preferences.";
 RL Biochem. J. 260:509-516(1989).
 [2]
 RN SEQUENCE OF 149-1225 FROM N.A.
 RX MEDLINE=89386015; PubMed=2780304;
 RA Janeczko R.A., Ramirez F.;
 RT "Nucleotide and amino acid sequences of the entire human alpha 1
 RT (III) collagen.";
 RL Nucleic Acids Res. 17:6742-6742(1989).
 [3]
 RN SEQUENCE OF 168-398.
 RX MEDLINE=77134724; PubMed=557335;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of cyanogen
 RT bromide peptides from the amino-terminal segment of type III collagen
 RT of human liver.";
 RL Biochemistry 16:1158-1164(1977).
 [4]
 RN SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 [5]
 RN REVIEW ON VARIANTS.
 RP Seyer J.M.;
 RL Submitted (DEC-1977) to the PIR data bank.
 [6]
 RN SEQUENCE OF 399-727.
 RX MEDLINE=79000343; PubMed=687591;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of five
 RT consecutive CNBr peptides from type III collagen of human liver.";
 RL Biochemistry 17:3404-3411(1978).
 [7]
 RN SEQUENCE OF 728-964.
 RX MEDLINE=80198282; PubMed=6246925;
 RA Seyer J.M., Mainardi C., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha 1
 RT (III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 19:1583-1589(1980).
 [8]
 RN SEQUENCE OF 950-1466 FROM N.A.
 RX MEDLINE=88189827; PubMed=3357782;
 RA Mankoo B.S., Daigleish R.;
 RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
 RL Nucleic Acids Res. 16:2337-2337(1988).
 [9]
 RN REVISION TO 1184.
 RP MEDLINE=89098346; PubMed=3211760;
 RX Molyneux K., Daigleish R.;
 RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
 RL Nucleic Acids Res. 16:11833-11833(1988).
 [10]
 RN SEQUENCE OF 1065-1466 FROM N.A.
 RX MEDLINE=85087944; PubMed=6096827;
 RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
 RA Rosenbloom J., Myers J.C.;
 RT "Molecular cloning and carboxyl-propeptide analysis of human type III
 RT procollagen.";
 RL Nucleic Acids Res. 12:9383-9394(1984).
 [11]
 RN SEQUENCE OF 965-1200.
 RX MEDLINE=81208139; PubMed=7016180;
 RA Seyer J.M., Kang A.H.;
 RT "Covalent structure of collagen: amino acid sequence of alpha
 RT 1(III)-CB5 from type III collagen of human liver.";
 RL Biochemistry 20:2621-2627(1981).
 [12]
 RN SEQUENCE OF 1176-1466 FROM N.A.
 RX MEDLINE=85157600; PubMed=2579949;
 RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
 RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
 RT (III) collagen. Partial characterization of the 3' end region of the
 RT gene.";
 RL J. Biol. Chem. 260:4357-4363(1985).
 [13]
 RN SEQUENCE OF 1161-1200 FROM N.A.
 RX MEDLINE=86187804; PubMed=3754462;
 RA Miskulin M., Daigleish R., Klue-Beckerman B., Rennard S.I.,
 RA Tolstoshev P., Brantly M., Crystal R.G.;
 RT "Human type III collagen gene expression is coordinately modulated
 RT with the type I collagen genes during fibroblast growth.";
 RL Biochemistry 25:1408-1413(1986).
 [14]
 RN SEQUENCE OF 1-170 FROM N.A.
 RX TISSUE=Placenta;
 RL MEDLINE=88303360; PubMed=3405773;
 RA Toman D., Ricca G., de Crombrughe B.;
 RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
 RT of human prepro alpha 1(III) collagen.";
 RL Nucleic Acids Res. 16:7201-7201(1988).
 [15]
 RN SEQUENCE OF 1-176 FROM N.A.
 RX MEDLINE=89378752; PubMed=2777083;
 RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
 RT "Cloning and analysis of the 5' portion of the human type-III
 RT procollagen gene (COL3A1).";
 RL Gene 78:255-265(1989).
 [16]
 RN REVIEW ON VARIANTS.
 RP MEDLINE=97255959; PubMed=9101290;
 RA Kuivaniemi H., Tromp G., Prockop D.J.;
 RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
 RT associated collagen (type IX), and network-forming collagen (type X)
 RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
 RL Hum. Mutat. 9:300-315(1997).
 [17]
 RN VARIANT AORTIC ANEURYSM ARG-303, AND VARIANT THR-668.
 RX MEDLINE=93293988; PubMed=8514866;
 RA Tromp G., Wu Y., Prockop D.J., Madhathi S.L., Kleiner C.,
 RA Earley J.J., Zhuang J., Noerregaard O., Darling R.C., Abbott W.M.,
 RA Cole C.W., Jaakkola P., Rymnen M., Pearce W.H., Yao J.S.T.,
 RA Majamaa K., Smullens S.N., Gatalica Z., Ferrell R.E., Jimenez S.A.,
 RA Jackson C.E., Michels V., Kaye M., Kuivaniemi H.;
 RT "Sequencing of cDNA from 50 unrelated patients reveals that mutations
 RT in the triple-helical domain of type III procollagen are an
 RT infrequent cause of aortic aneurysms.";
 RL J. Clin. Invest. 91:2539-2545(1993).
 [18]
 RN VARIANT THR-698.
 RX MEDLINE=91045136; PubMed=2235526;
 RA Zafarullah K., Kleiner C., Tromp G., Kuivaniemi H., Kontusaari S.,
 RA Wu Y., Ganguly A., Prockop D.J.;
 RT "G to A polymorphism in exon 31 of the COL3A1 gene.";
 RL Nucleic Acids Res. 18:6180-6180(1990).
 [19]
 RN VARIANT AORTIC ANEURYSM ARG-786.
 RX MEDLINE=91056145; PubMed=2243125;
 RA Kontusaari S., Tromp G., Kuivaniemi H., Romanic A.M., Prockop D.J.;
 RT "A mutation in the gene for type III procollagen (COL3A1) in a family

with aortic aneurysms.", J. Clin. Invest. 86:1465-1473 (1990).	856	GKAEPPVKSECTEEAEAGPAKGDAAEAATAEGALKAEKKGSGRATTAKSSGAPQDS	911
[19]	221	-----IGPSGA-GKGE-----SRP----	236
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	916	DSSATCSADEVDEAEAGDKNNLLSPRSLTPTGDP-ANASPOKPLDLKOLKORAAAI	974
[20]	237	-----GRPGERGLPPGIGKPGAGIPGFGMGKGRGFGDNGEKGTGAP	281
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	975	PIOVTKVHEPPREDAAATKAPPAPPPONLQPSDAPQPGSS-----PRGKSRPA	1021
[21]	282	GLK-----GKGLPGENGAPGMPGRCAGRCGRGRCPLPGAGRGNDGARGDGGQ	333
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1028	PPADKAEFAAEAKLPGDPCWTSGLPFPVPPREVIKASPHAPDPSAFSYPAGHPLFLG	1081
[22]	334	PPG-----PPGTAG--FPGSPKAGEVGPAGSPGNGAPGQRPFGQ	375
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1088	LHDTARVLPRLPTTISNPPPLISSAKHPSVLEROIKAISQGMVQLHVPYSEHAKAPVP	1147
[23]	376	HAGAQP--PGPPGINSFG-----GKGMGPAGIPGAGLGMARGKGGPPG	418
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1148	V-TMGLP-----LPMDFPKLAPFSGVKQEQLSPRGAGPPESLGVTAQASVLRGT	1198
[24]	419	AGANGAPGLRGAGBPCK-----NGAKGEP-GRGERGEAGIPGVKAGKEDKDGSPGE	473
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1199	ALGSPVGGSTIKGIPSTRVPSDSAITYRGSIHTGTPADVLKGTITRIIGEDSRSLDRG	1258
[25]	474	GANGLPGAAGERGAGFRGPAG-----PNGIP-----GKGPAG-ERG	510
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1259	REDSL-PKGHVIVEGKKGHLVSEYGMSTVQCKEDGRSS-----GPPHETAAPKRTYDM	1313
[26]	511	APGAPGRGAAGEPGRDG-----VPGPGMRGMPGSPGPGSDGRGPPGS	556
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1314	MEGRVCR-----AISASIEGLMGRAIIPERH-SPHLKQKHRRIGSTIQGIPRSYVEAQ	1367
[27]	557	-QGESGRPGPPGPGSPRGQPGVMGPPGPKGNDGAFGKNGERGPGGPGQPGPKNGBTG	615
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1368	EDYLREAKLLKREGTPPPPPSRDLTAYKTAQALGPLKLPAPHEGLVATVKEAGRSIHE	1427
[28]	616	P-----QPPPGTGGGDKGT-----GP-----PGQGLQLPCTGCP	649
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1428	IPRELRHTPELPAPRLKEGSIITQGTPLKYDTGASTGSKGHVRSILIGSPQRTPPV	1487
[29]	650	-PENGK-----PGEFPG-----KGDAGAPGAPGCKGDA-----CAPGERGPP	686
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1488	HLDVMDARALACRYEESLKRPGTSSSGSISARCAPVIVPELKGPRSPITYEDHG	1547
[30]	687	-----GLAGAPL-----RGGA-----GPP-----GEGGKGAAGP	714
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1548	APFAGHLPRGSPVTWREPTPLQEGSLSSKASQDRKLTSTPREIAKSPHSTVPEHHHP	1607
[31]	715	PPGAA-----GTPGLQMPGER-----GGLGSPGPKGDKGPPGPG	750
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1608	ISPYEHLRGVSGVDLYRSHIPLAFDPTSIPRGIPLDAAYLYLPRHLAPNPTPHLYPP	1667
[32]	751	-----ADGVEPKD-----GPRG-----PTGP-IGPP	770
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1668	YLIRGYPTAALLENQTIINDYITTSQQMHNTATAMARADMLRGLSPRESSLALNAA	1727
[33]	771	-----GPGQPGDKGEGAP-----GLPGIAG	792
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1728	PRGIIDLSQVPHLPVLPPTGCTPA-TAMDRLAYLTPAT-----OPFSRHSISP	1780
[34]	793	PPG-----SPGERGTGPPGAPGQNGEPGKGERGAPGEGEG	836
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1781	GPHTLTKPTTSSSRERDRDRDREREKSLTSTTTVEHAPIWRPGTEQSSSGSSGS	1840
[35]	837	GPPGVAGPPGS-----GPGPPGQGVKGERGS	865
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1841	SCGGG-----GSSSRPASHAHQHSPLSPPTQDALQORPSVLHNTGMKIIITAVEPSKP	1895
[36]	866	PGCGAAGFPAGLUPGPPGNGPNPSPGSGPKDGPFGPAGNTGAP	915
RICHARDS A.J., Narcisi P., Lloyd J.C., Ferguson C., Pope F.M.; "The substitution of glycine 661 by arginine in type III collagen produces mutant molecules with different thermal stabilities and causes Ehlers-Danlos syndrome type IV.", J. Med. Genet. 30:690-693 (1993).	1896	TVLRSTSTSSPV-----RAAATPPATHPLGTLGDVVYTLMEVLLPKAPRV	1945

alpha4(IV) collagen chains are arranged head-to-head on chromosome 2q36.";
 FEBS Lett. 424:11-16(1998).
 [3]
 RP SEQUENCE OF 1219-1690 FROM N.A.
 RC TISSUE-EYE; PubMed=8365481;
 RX MEDLINE=93374047; PubMed=8365481;
 RA Sugimoto M., Ohashi T., Yoshioka H., Matsuo N., Ninomiya Y.;
 "cDNA isolation and partial gene structure of the human alpha 4(IV)
 collagen chain.";
 FEBS Lett. 330:122-128(1993).
 [4]
 RP SEQUENCE OF 1407-1507 FROM N.A.
 RX MEDLINE=93054733; PubMed=1429714;
 RA Kamagata Y., Mattei M.-G., Ninomiya Y.;
 "Isolation and sequencing of cDNAs and genomic DNAs encoding the
 alpha 4 chain of basement membrane collagen type IV and assignment of
 the gene to the discal long arm of human chromosome 2.";
 J. Biol. Chem. 267:23753-23758(1992).
 [5]
 RP REVIEW ON VARIANTS.
 RX MEDLINE=97338662; PubMed=9195222;
 RA Lemmink H.H., Schroeder C.H., Monnens L.A.H., Smeets H.J.M.;
 "The clinical spectrum of type IV collagen mutations.";
 Hum. Mutat. 9:477-499(1997).
 [6]
 RP VARIANT AS SER-1201.
 RX MEDLINE=95078927; PubMed=7987396;
 RA Mochizuki T., Lemmink H.H., Mariyama M., Antignac C., Gubler M.-C.,
 Pirson Y., Verellen-Dumoulin C., Chan B., Schroeder C.H.,
 Smeets H.J.M., Reenders S.T.;
 "Identification of mutations in the alpha 3(IV) and alpha 4(IV)
 collagen genes in autosomal recessive Alport syndrome.";
 Nat. Genet. 8:77-82(1994).
 [7]
 RP VARIANT FBH GLU-897.
 RX MEDLINE=96379660; PubMed=8787673;
 RA Lemmink H.H., Nillesen W.N., Mochizuki T., Schroeder C.H.,
 Brunner H.G., van Oost B.A., Monnens L.A.H., Smeets H.J.M.;
 "Benign familial hematuria due to mutation of the type IV collagen
 alpha4 gene.";
 J. Clin. Invest. 98:1114-1118(1996).
 [8]
 RP VARIANTS AS, AND VARIANTS.
 RX MEDLINE=99011253; PubMed=9792860;
 RA Boye E., Mollet G., Forestier L., Cohen-Solal L., Heidet L.,
 Cochat P., Gruenfeld J.-P., Palcoux J.-B., Gubler M.-C., Antignac C.;
 "Determination of the genomic structure of the COL4A4 gene and of
 novel mutations causing autosomal recessive Alport syndrome.";
 Am. J. Hum. Genet. 63:1329-1340(1998).
 CC -!- FUNCTION: Type IV collagen is the major structural component of
 glomerular basement membranes (GBM), forming a 'chicken-wire'
 meshwork together with laminins, proteoglycans and entactin/
 nidogen.
 CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
 alpha 6(IV), each of which can form a triple helix structure with
 2 other chains to generate type IV collagen network.
 CC -!- SUBCELLULAR LOCATION: Cell surface (Potential).
 CC -!- TISSUE SPECIFICITY: Alpha 3 and alpha 4 type IV collagens are
 colocalized and present only in basement membranes of kidney, eye,
 cochlea, lung and brain.
 CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
 domain (NC1) at their C-terminus, frequent interruptions of the G-
 X-Y repeats in the long central triple-helical domain (which may
 cause flexibility in the triple helix), and a short N-terminal
 triple-helical 7S domain.
 CC -!- PTM: Prolines at the third position of the tripeptide repeating
 unit (G-X-Y) are hydroxylated in some or all of the chains.
 CC -!- PTM: Type IV collagens contain numerous cysteine residues which
 are involved in inter- and intramolecular disulfide bonding. 12 of
 these, located in the NC1 domain, are conserved in all known type
 IV collagens.
 CC -!- DISEASE: Defects in COL4A4 are a cause of autosomal recessive

CC Alport syndrome (AS) [MIM:203780], an hereditary disorder
 characterized by progressive glomerulonephritis, renal failure,
 hematuria, ocular abnormalities and deafness. The recessive form
 occurs equally between males and females.
 CC -!- DISEASE: Defects in COL4A4 are a cause of familial benign
 hematuria (FBH) [MIM:141200] or thin basement membrane disease.
 CC FBH is characterized by persistent hematuria, an electron
 microscopically detectable thin glomerular basement membrane (GBM)
 and an autosomal dominant mode of inheritance. Renal function
 remains normal. In children, differentiation between FBH and AS
 can be difficult, because both disorders are manifested by
 persistent hematuria and thin GBM at that age.
 CC -!- SIMILARITY: TO OTHER TYPE IV COLLAGENS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X81053; CAA56943.1; -;
 DR EMBL; AB008496; BAA25065.1; -;
 DR EMBL; D17391; BAA04214.1; -;
 DR PIR; A55360; CGHUIB.
 DR Gene; HGNC:2206; COL4A4.
 DR MIM; 120131; -;
 DR MIM; 141200; -;
 DR MIM; 203780; -;
 DR InterPro; IPR008161; C1g_helix.
 DR InterPro; IPR008160; Collagen.
 DR InterPro; IPR014442; Procollagn4_C.
 DR Pfam; PF01413; C4; 2.
 DR Pfam; PF01391; Collagen; 21.
 DR ProDom; PD000007; C1g_helix; 3.
 DR ProDom; PD003923; ProcollagnC4; 1.
 DR SMART; SM00111; C4; 2.
 KW Extracellular matrix; Connective tissue; Basement membrane; Repeat;
 KW Hydroxylation; Collagen; Glycoprotein; Signal; Disease mutation;
 KW Polymorphism; Alport syndrome.
 FT SIGNAL 1 38 POTENTIAL.
 FT CHAIN 39 1690 COLLAGEN ALPHA 4(IV) CHAIN.
 FT DOMAIN 39 64 7S DOMAIN.
 FT DOMAIN 65 1459 TRIPLE-HELICAL REGION.
 FT DOMAIN 1460 1690 NONHELICAL REGION (NC1).
 FT SITE 94 96 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 145 147 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 189 191 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 310 312 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 724 726 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 785 787 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 989 991 CELL ATTACHMENT SITE (POTENTIAL).
 FT SITE 1206 1207 CLEAVAGE (BY COLLAGENASE)
 (BY SIMILARITY).
 FT SITE 1212 1214 CELL ATTACHMENT SITE (POTENTIAL).
 FT DISULFID 1480 1569 OR 1566 (BY SIMILARITY).
 FT DISULFID 1513 1566 OR 1569 (BY SIMILARITY).
 FT DISULFID 1525 1531 BY SIMILARITY.
 FT DISULFID 1588 1686 OR 1683 (BY SIMILARITY).
 FT DISULFID 1622 1683 OR 1686 (BY SIMILARITY).
 FT DISULFID 1634 1641 BY SIMILARITY.
 FT CARBOHYD 142 142 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 569 569 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 441 446 Missing (in AS).
 FT VARIANT 545 545 /FTId=VAR_008148.
 FT VARIANT 570 570 G -> A (in dbSNP:1800516).
 FT VARIANT 897 897 E -> Q.
 FT VARIANT 897 897 /FTId=VAR_008150.
 FT VARIANT 931 931 G -> E (in FBH).
 FT VARIANT 931 931 /FTId=VAR_001912.
 FT A -> T.

FT	CARBOHYD	930	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1277	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1293	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1557	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	1775	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2192	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	2721	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3419	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	3948	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4745	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4901	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4958	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4965	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	4987	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5037	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5052	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5156	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5427	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5467	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5506	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5507	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5543	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5553	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5604	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5618	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	5662	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	VARIANT	5137	T -> S (in dbSNP:2672788).	
FT	CONFLICT	34	/FTid=VAR 014123.	
FT	CONFLICT	95	G -> E (IN REF. 2).	
FT	CONFLICT	104	FPGLCN -> LPLCK (IN REF. 2).	
FT	CONFLICT	142	S -> C (IN REF. 2).	
FT	CONFLICT	142	E -> K (IN REF. 1).	
FT	CONFLICT	225	R -> S (IN REF. 2).	
FT	CONFLICT	330	PL -> T (IN REF. 2).	
FT	CONFLICT	337	E -> N (IN REF. 2).	
FT	CONFLICT	356	E -> K (IN REF. 2).	
FT	CONFLICT	362	G -> R (IN REF. 2).	
FT	CONFLICT	369	MISSING (IN REF. 2 AND 3).	
FT	CONFLICT	374	D -> N (IN REF. 2).	
FT	CONFLICT	393	RT -> TR (IN REF. 2).	
FT	CONFLICT	468	RK -> GR (IN REF. 2).	
FT	CONFLICT	512	L -> P (IN REF. 2).	
FT	CONFLICT	585	GAA -> AH (IN REF. 3).	
FT	CONFLICT	601	A -> S (IN REF. 3).	
FT	CONFLICT	628	DP -> RS (IN REF. 2).	
FT	CONFLICT	633	F -> L (IN REF. 2).	
FT	CONFLICT	676	A -> P (IN REF. 3).	
Query Match 2.5%; Score 335.5; DB 1; Length 5703;				
Best Local Similarity 19.5%; Pred. No. 0.0013;				
Matches 409; Conservative 203; Mismatches 835; Indels 649; Gaps 91;				
Qy	687	PAASEEAAFPVVEDEMEASGVSGNEEVEAEALHASGNEVPRGECSPATVNSS	746	
Db	2149	PSSTPGTPIPPVLTTATTAAATSSST-----VTPSSAL-----GTTHTPVPVNTTA	2195	
Qy	747	DTE-----SISPH-----TEAKDTQNGKPKPATLGA---DGPPP	780	
Db	2196	TTHGRSLSPSPSHVCTAWTSATSGILGTTHITEPSTGTSHT---PAATTGTTQHSPTAL	2252	
Qy	781	GPPTPRRTSRAPTEP-----TPASEATGATPP-----PAPPSPSAPPPVVPK	824	
Db	2253	SSPHSPRTTESPPSPGTTTTPGHTTATSRITATATSKTRTSILLPSOPT-SAPITTVT	2311	
Qy	825	EKEEETAAAPVVEEGEQKPPAAEEL-----AVDTGKAEPVKSECTEEAEEGPAK	877	
Db	2312	MGCEPQCAWSEWLDYSYMPGPGSGGDFDYSNIRAAGGACVCEQPLGLECRAQOPGVPLR	2371	
Qy	878	KDAEAEATAEAGALKAEKKEG-----GSGRATTAKSSGAPQSDSS	918	
Db	2372	ELQGVVECSLDFGLVCRNRQGVKFKMCFNVEIRVFCNCVYCHCPSTATSTATSPSTPG	2431	
Qy	919	ATCSADEVDEAEGDKNRLSLPRPSLLTPTGDPNANASPOKPLDLKQKRAAIIPIQV	978	

Db	2432	TTW-----ILTELTTATTTESTGSTATPTSTL-----RTAPPPKULT	2469	
Qy	979	TKVHEPPREDAAPYKPAAPPQPNQLOPSDAPQOQPGSPGRKSRSPAPADKEAPAAE	1038	
Db	2470	TTATTP---TVTSSKATPSSSPGTATALP---ALRSTATTATSVTPIPSSSLGT---	2519	
Qy	1039	AQKLPDGPDPQWTSGLPFVPVPREVIKASPHAPDPSAF-----SYAPPGH	1082	
Db	2520	-----TWRLSQTTTPTATMSTATPSTETAHTSVLTATATTGATGSAVATPS-	2569	
Qy	1083	PLPLGLHDTARPVL-----PREPTISNPPPLISSAKHPSVLERQICAISSQGNMSV-	1131	
Db	2570	STPGTAHTTKVPTITTTTGTATPSSSPGTALTTPVWLSTTTTPT-----TRGSTVT	2620	
Qy	1132	QLHVPYSEHAKAPGVPTMGLPLPMDPKLAPSGVKQEQOLSPRGQAGPPESIGVPTAQE	1191	
Db	2621	PSSIPGTHHTATVLTITTTTVA-----TGSMATPSSSTQTSCTGTPSSL---TTTA	2666	
Qy	1192	ASVLRGTAIGSVPGSGITKGI-PSTRVPSDALTIRGSIITHGTPADVLVYKGTITRIIGED	1250	
Db	2667	TTI---TATGSTNPSSTPGTRPIPPVLTTATTAPATSTVTPSSAL--GT-THTPPVP	2720	
Qy	1251	SPRLDRGREDSLPKGHVIYEGKGHVLSYEGGMSVTQCSKEDGRSSSGPPHETAAPKRT	1310	
Db	2721	NTTATTGRLSLSPSPSHV---RTAWTSATSGTLGTHITE---PSTGTSHTPAATTGT	2773	
Qy	1311	YDMMEGRVGRAISSASIEGLMGRAIP-PERHSHPLHKEQHIIIRGSIQTGIPRYSVYEAQED	1369	
Db	2774	TQ-----HSTPALSPHPSSTRTTESPPSPGTTTTPGH-----TTATSRTTATATPS	2818	
Qy	1370	YLREAKLLKREGTPPPPPSRDLTEA-----YKQALGLPLKPAHEG	1413	
Db	2819	KRTSTILL-----PSSPTSABITTVTMGCPQCAWSEWLDYSVMPGP---SGDFFD	2868	
Qy	1414	LVAIVKEAGRSIHEIPEELRHHTPELAPR-----PLKEGSIQTGTPLYKYDTGASTT	1466	
Db	2869	YSNIRAAGGAVCEQ-----LGLCRAQAQGVPLRE-LQGVVECSLDFGLVCR	2917	
Qy	1467	GSKKHVDVRSILGSPGRTFFPVPHPLDVMDARALERACYEESLK-----SRGTAS	1516	
Db	2918	N-----REQVGK-----FKMCFNVEIRVFCNCYCHCPSTPATSS	2951	
Qy	1517	SSGGSIAAGAPVIVPELKGKPROSLTYEDHGAFFA--GHLPRGSPVTMRPTPRLQEGSL	1574	
Db	2952	TATPSSTPGTTWLITQTAAATTTATTGTAIPSSTPGTAPPKVLTSQATTP-----TA	3006	
Qy	1575	SSKA---SODRKLSTPREIAKSPHSTVPEHHHPHPISPYEHLLRGVSGVDLYRSHIPLA	1631	
Db	3007	TSSKATSSSPRTATTLPLVTSTATKSTATSTFTPIPSSTL-----GTTGTSQNRPPHPMA	3061	
Qy	1632	FDPTSIPRGIPLDAAAAYVLPRLHAPNPTYPHLYPPYLRGVPDTAALNRQTIINDYIT	1691	
Db	3062	TMSI-----IHPSST-----PETT-----	3075	
Qy	1692	SOOMHENTATAMAQRAADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLPPTPGTP	1751	
Db	3076	-----HTSVLTTKATTTTATSSMSTPSS-----TPGIT	3104	
Qy	1752	-----ATAMDLRLAYLPTAQPFSSSHSSPLSPGPGTHLTPTTTTSSSERDRDRRED	1805	
Db	3105	WILTELTATTAATTAALP-----HGTFPSSTPGTWTILTEPSTTATTVTPTGTATAS	3155	
Qy	1806	ROREREKS--ILTSTTTTVEHAPIWRPGTQSGSGSGSGSGSGSSRRPASHAHQHSPI	1863	
Db	3156	SIFATAGTLKVLSTATTP-----TVISBRATPSS--SPGTATPALPALSTA-----	3200	
Qy	1864	SPRTDALOQRPVSLHNTCMKGIITAVEPSKPTVLRSTSTSSPVRPAATFPFATHCPLGG	1923	
Db	3201	TTPTATSVTAIPSSSLGTAWTRLSTQTTTP-----TATMSTATPSSST-PETVHTSTVL	3251	
Qy	1924	TLDGVVPTLMEVLLPKAPRVARPERPRADTGHAFKAPPPARSG-----LEPASSP	1975	

Db 3252 TTTAT-TTRTGSVATPSSTPGTAHTTKVPTTTTTTGTATPSSPGTALTTPPVWISTTTTTP 3310
Qy 1976 -SKGSEPRPLVPVSGH-ATIAATPAKNAHPHASPDPAPPAGASDPHREKTSKPPSI 2033
Db 3311 TTRGSTVPISPGTHATVLTITTTTAVATGSMATPSSSTQTSPTSLTTAT----- 3365
Qy 2034 QELBRLSLGVHGSYSPE--GVFVSPV-----SSPSLTHDKGLPKHLEBLDKSHLEGE 2085
Db 3366 ----TITATGTTNPSSTPGTTPIPVLTITTTATPAATSTVTPS--SALGTTH----- 3413
Qy 2086 LRPQGPVKLGGEAAHLPHLPBPESOPSSPILLOTA--PGVKGHQVWTLAQHISEVI 2143
Db 3414 -TPVPV-----NTTATHGSRPLPPSPHTVPTAWTSAATSGILG-----TTHITS-- 3456
Qy 2144 TQDTRHHQPQSLAPLAPLAPLAPLPPSDLVLDLRPPDLVLPDPDHGAPARGSPHSGG 2203
Db 3457 -----PSTGTSHTPAATGTTPQSTPALSSHPSPRTTESPPSGTTTPG--HTGRT 3506
Qy 2204 KR-----SPENKTSVL-----GGGED 2220
Db 3507 SRTTATATPSKTRTSTLLPSSPTSAPITTVVTGCEPOCANSEWLDVSYMPGPGSGDFD 3566
Qy 2221 -----CIE-----PVSPEGTEPECH--SRSAVYPLLVRDGEQ----- 2251
Db 3567 TYSNIRAAGVACQPLGLECRAQAQPGVPLRELQGVVECSLDPLGVCNRNEQVKGKFKMC 3626
Qy 2252 -----TPSRMGSKSPGNTSQPPAFFSKLT-----ESNSAMVKSQKQ 2288
Db 3627 FNYEIRVECCNYGCHPSTPATSTSTGTTGTTWILTKLTITTTATTESTGSTATPSTQ 3686
Qy 2289 -----EINKKLTHNRNPEYNIQFTEIFNMPAITGTLMTYRSQVQEHAS 2337
Db 3687 GPPAGTHVSTATTPTVTSKATP---FSFSG-----TATAPALRSTATTPTAT 3734
Qy 2338 TNWGLEAIRKALMGKYDQWESPLSANAPNLNASASLPAAMPITAADGRSHT---L 2394
Db 3735 S---FTAPSSSLGTTWR-----LSQTTTPMATMSTATPSSPTETV-----HTSTVL 3779
Qy 2395 TSPGGGKAKYGRPSSRKAKSPAGLASGRPPSVSS-----VHSEGDGNRRTPLTNRV 2449
Db 3780 TT-----TATTGATGVSATPSTPGTAHTTKVPTTTTTTGTVPSSPGTARTP---PV 3831
Qy 2450 W---EDRPSAGSTPFP-----YNPLIMR-----LOAGVMASPPPPGLPAGSGP 2490
Db 3832 WISITTTPTSGSTVPSSIPGTHTPVLTITTPVLTITTPQVATGSMATPSSSTQTSSTGTP 3887
RESULT 71
CALH HUMAN
ID CALH HUMAN STANDARD; PRT; 1516 AA.
AC P39050; Q9UK38; Q9Y6Q7; Q9Y6Q8;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Collagen alpha 1(XVIII) chain precursor [Contains: Endostatin].
GN COL18A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=98164096; PubMed=9503365;
RA Saarela J., Vlika R., Rehn M., Purmonen S., Pihlajaniemi T.;
RT "Complete primary structure of two variant forms of human type XVIII
RT collagen and tissue-specific differences in the expression of the
RT corresponding transcripts."
RL Matrix Biol. 16:319-328(1998).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=20289799; PubMed=10830953;
RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,

Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA Minoshima S., Shimizu N., Nordsiek G., Hornisch K., Brandt P.,
RA Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA Ramser J., Beck A., Klages S., Hennig S., Riessmann L., Dagand E.,
RA Wehrmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,
RA Lebrach H., Reinhardt R., Yaspo M.-L.;
RT "The DNA sequence of human chromosome 21."
RL Nature 405:311-319(2000).
RN (3)
RP SEQUENCE OF 834-1516 FROM N.A.
RX MEDLINE=94245237; PubMed=8188291;
RA Oh S.P., Warman M.L., Seldin M.F., Cheng S., Knoll J.H., Timmons S.,
RA Olsen B.R.;
RT "Cloning of cDNA and genomic DNA encoding human type XVIII collagen
RT and localization of the alpha 1(XVIII) collagen gene to mouse
RT chromosome 10 and human chromosome 21."
RL Genomics 19:494-499(1994).
RN (4)
RP SEQUENCE OF 1334-1516 FROM N.A.
RC TISSUE=Placenta;
RA Zhi-Yong H., Biao L., Wei-Jie Z., Xiang-Fu W.;
RT "Cloning and expression of human endostatin gene in Escherichia
RT coli."
RN Submitted (SEP-1999) to the ENBL/GenBank/DBAJ databases.
RN (5)
RP INVOLVEMENT IN KNOBLOCH SYNDROME.
RX MEDLINE=20400145; PubMed=10942434;
RA Sertie A.L., Sossi V., Camargo A.A., Zatz M., Brahe C.,
RA Passos-Bueno M.R.;
RT "Collagen XVIII, containing an endogenous inhibitor of angiogenesis
RT and tumor growth, plays a critical role in the maintenance of retinal
RT structure and in neural tube closure."
RL Hum. Mol. Genet. 9:2051-2058(2000).
RN (6)
RP VARIANT ASN-1437.
RX MEDLINE=21518361; PubMed=11606364;
RA Iughetti P., Suzuki O., Godoi P.H., Alves V.A., Sertie A.L.,
RA Zorick T., Soares F., Camargo A., Moreira E.S., di Loreto C.,
RA Moreira-Filho C.A., Simpson A., Oliva G., Passos-Bueno M.R.;
RT "A polymorphism in endostatin, an angiogenesis inhibitor, predisposes
RT for the development of prostatic adenocarcinoma."
RL Cancer Res. 61:7375-7378(2001).
CC -!- FUNCTION: COL18A1 probably plays a major role in determining the
CC retinal structure as well as in the closure of the neural tube.
CC -!- FUNCTION: Endostatin potentially inhibits endothelial cell
CC proliferation and angiogenesis. May inhibit angiogenesis by
CC binding to the heparan sulphate proteoglycans involved in growth
CC factor signalling.
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Long; Synonyms=NC-493;
CC IsoId=P39060-1; Sequence=Displayed;
CC Name=Short; Synonyms=NCL-303;
CC IsoId=P39060-2; Sequence=VSP_001155, VSP_001156;
CC -!- TISSUE SPECIFICITY: Present in multiple organs with highest levels
CC in liver, lung and kidney.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- POLYMORPHISM: There is an association between a polymorphism in
CC position 1437 and prostate cancer. Heterozygous Asn-1437
CC individuals have a 2.5 times increased chance of developing
CC prostate cancer as compared with homozygous Asp-1437 individuals.
CC -!- DISEASE: Defects in COL18A1 are a cause of Knobloch syndrome (KNO
CC [MIM:267750]); an autosomal recessive disorder defined by the
CC occurrence of high myopia, vitreoretinal degeneration with retinal
CC detachment, macular abnormalities and occipital encephalocele.
CC -!- SIMILARITY: BELONGS TO THE FIBRIL-ASSOCIATED COLLAGENS WITH
CC INTERRUPTED HELICES (FACIT) FAMILY.
CC -!- SIMILARITY: Contains 1 TSP N-terminal (TSPN) domain.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

 EMBL; AF018081; AAC39658.1; -
 EMBL; AF018082; AAC39659.1; -
 EMBL; AL163302; CAB90482.1; -
 EMBL; L22548; AAC51864.1; -
 EMBL; AF184060; AAF01310.1; ALT_INIT.
 PDB; 1BNL; 02-DEC-98.
 GlycoSuiteDB; P39060; -
 Genew; HGNC:2195; COL18A1.
 MIM; 120328; -
 MIM; 267750; -
 GO; GO:0005581; C:collagen; TAS.
 GO; GO:0007397; P:histogenesis and organogenesis; TAS.
 GO; GO:0008285; P:negative regulation of cell proliferation; TAS.
 GO; GO:0007601; P:vision; TAS.
 InterPro; IPR008161; Clg_helix.
 InterPro; IPR008160; Collagen.
 InterPro; IPR008985; ConA like lec_gl.
 InterPro; IPR001791; Laminin_G.
 Pfam; PF01391; Collagen; 7.
 Pfam; PF02210; TSPN; 1.
 ProDom; PD000007; Clg_helix; 1.
 SMART; SM00282; LamG; 1.
 SMART; SM00210; TSPN; 1.
 Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 Cell adhesion; Collagen; Glycoprotein; Signal; Alternative splicing;
 Polymorphism; 3D-structure.
 SIGNAL 1 23
 FT CHAIN 24 1516
 FT CHAIN 1334 1516
 FT DOMAIN 221 409
 FT DOMAIN 410 516
 FT DOMAIN 517 550
 FT DOMAIN 551 560
 FT DOMAIN 561 640
 FT DOMAIN 641 664
 FT DOMAIN 665 786
 FT DOMAIN 787 809
 FT DOMAIN 810 892
 FT DOMAIN 893 906
 FT DOMAIN 907 948
 FT DOMAIN 949 961
 FT DOMAIN 962 1034
 FT DOMAIN 1035 1044
 FT DOMAIN 1045 1077
 FT DOMAIN 1078 1089
 FT DOMAIN 1090 1111
 FT DOMAIN 1112 1118
 FT DOMAIN 1119 1173
 FT DOMAIN 1174 1186
 FT DOMAIN 1187 1204
 FT DOMAIN 1205 1516
 FT CARBOHYD 68 68
 FT CARBOHYD 129 129
 FT CARBOHYD 164 164
 FT CARBOHYD 591 691
 FT CARBOHYD 1329 1329
 DISULFID 1366 1506
 DISULFID 1468 1498
 SITE 1095 1097
 VARSPLIC 1 180
 VARSPLIC 181 215

PRCPWPRRRRLDLVLAFLVLLGVRAASAP (in isoform Short).
 /FTid=VSP_001156.
 D -> N (increased risk of developing prostate cancer).
 /FTid=VAR_012709.
 F -> S (IN REF. 2).
 I -> V (IN REF. 2).
 V -> L (IN REF. 3).
 P -> R (IN REF. 3).
 P -> L (IN REF. 3).
 P -> L (IN REF. 3).
 A -> P (IN REF. 3).
 L -> K (IN REF. 3).
 P -> A (IN REF. 3).
 P -> PGP (IN REF. 2).
 G -> GO (IN REF. 3).
 R -> G (IN REF. 3).
 A -> G (IN REF. 3).
 LR -> CG (IN REF. 3).
 R -> T (IN REF. 4).
 S -> Y (IN REF. 4).
 SQ SEQUENCE 1516 AA; 153840 MW; 3C70F29A4476EE76 CRC64;
 Query Match 2.5%; Score 335; DB 1; Length 1516;
 Best Local Similarity 22.8%; Pred. No. 0.0004;
 Matches 304; Conservative 89; Mismatches 519; Indels 420; Gaps 71;
 QY 675 KERNARRKKKAPAAAEAAFPVVEDEMEASGVSGNEEEMVEAEALHASGNEVPRG 734
 DB 443 REETGAALKRPLAP-----PPVTPLAGSSSTEDSRSEVEEQTVAISLGAQT--- 492
 QY 735 ECGSPATVNNSDTESIPSHTEAAKDTGONG----PKPPATLGADGPP--PGPPTPPRR 788
 DB 493 -LFGSDSVSTWDGVSRTPGGRVKEGGLKGKGPVGPVGGPGRAGPGSPCLPGLPCP 551
 QY 789 TSRAPTEPT-PASEATGATPPAPSPAPPPVWVVEKEEHEATAAPVVEESEEOKPPA 847
 DB 552 VS--PLGPAGPALQTVPGPGPPGPRDGT-----GRDGE-----PGDGEDKPG- 597
 QY 848 AEELAVDTGKAEEFVKSECTEEAEAGPAKGDAAEAEATAEGALKAEKKEGSGRATTAK 907
 DB 598 -----DTGPQGF-----GTPGDVGP-----KDKGDPGVGERGPG 629
 QY 908 SSGAPQDSSS-----ATCSADEVDEAGDKNRLLSPRSLTPTGDPANASPKPLDL 963
 DB 630 PQGPPGPPGPFRRHDKLTFIDMEGSGFGDLEALGRF-GFPGPPGPPGVPLGPEP--- 685
 QY 964 KOLKQRAAIP-PIQVTKVHEPRE--DAAPTKPAPPAPP---PP-----QNLQPEDA 1011
 DB 686 GRGVNNSDVFGPAGLPGV--FGREGPPGPGPLGPPGPPGREGPPGRTGQKSLGEAGA 743
 QY 1012 PQPGS-----SPRKSRSPAPPA-----DKE----- 1033
 DB 744 PGHKGSKGAPGAPGARGESGLAGAPGAPGPPGPPGPPGPPGPPGPPGPPGPPGPPG 803
 QY 1034 -AFAAEAQK-----LPGDPPCWTSGLP-----FP-VPREVTKASHPAD 1071
 DB 804 TARSADGPPGPPGLPKRGDP--GVPLPGAKEGVADGIPGPPGLPREGI-AGPQPK 860
 QY 1072 PSAFSYAPPGHPLPLGLHDTARPLPRPTISNPPPLISSAKHPSVLERQIGAIQGMVS 1131
 DB 861 GDRGSRGKGDG--GKDGVGQPLGPP--GPPGPVV-----YVSEQDSV----- 902
 QY 1132 QLVHYSEHAKAPVG-PVTMGLPLPMDPKLAPFSGVQEQLS--PRGAGPESLGVPT 1188
 DB 903 -LSVP-----GPEGRPGFAGPPGAPKGNL---GSKGERGSPGKKEGEPGIFSPD 952
 QY 1189 AQEASVLRGTALGSPVGGSIITKIGPSTRVPSDAITYRGSITHGTADVLYKGTTRIIG 1248
 DB 953 -----GGALGPAQKA--KGEFGFRGPG-----PYGRPG--YKGEI----- 985

Db 1527 SGVAGTGVGVFNHVSASAHFPGARVGRGAACGATLGPSPLOAASASASSFOAPA 1586
Qy 1980 -----EPRLVPPVSGHATARTPAKNLAPHAGSPD-----PPAPPASADPHREKTQ 2027
Db 1587 SVETRPDP 1646
Qy 2028 SKPFSIQEILRSIGYHSSVSPGVSPVSPVSPVSPVSPVSPVSPVSPVSPVSPVSPV 2087
Db 1647 S-----LASL-----PPNQAFLPPTSASL-----1666
Qy 2088 PKQGPVKLGEEAAHPLRLPLPSQSPSSP-----LLQTAPGVKGHQRVVTLAQHISEV 2142
Db 1667 --PPANASLSIKTSLPHKGRSPSTVHHQPLRLALAAQAPGIPQASATGPSAVVWSLG 1724
Qy 2143 ITQDYTRH 2150
Db 1725 MPPPYAAH 1732

RESULT 73
SMF1_HUMAN
ID SMF1_HUMAN STANDARD; PRT; 1902 AA.
AC O14497; Q0UPZ1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE SWI/SNF-related, matrix-associated, actin-dependent regulator of
DE chromatin subfamily F member 1 (SWI-SNF complex protein p270) (B120).
GN SWARCF1 OR CLORF4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND MUTAGENESIS.
RX MEDLINE=20221560; PubMed=10757799;
RA Dallas P.B., Pacchione S., Wilsker D., Bowrin V., Kobayashi R.,
RA Moran E.;
RT "The human SWI-SNF complex protein p270 is an ARID family member with
RT non-sequence-specific DNA binding activity";
RL Mol. Cell. Biol. 20:3137-3146(2000).
RN [2]
RP SEQUENCE OF 1-1175 FROM N.A.
RX MEDLINE=98094256; PubMed=9434167;
RA Takeuchi T., Chen B.-K., Oiu Y., Sonobe H., Ohtsuki Y.;
RT "Molecular cloning and expression of a novel human cDNA containing CAG
RT repeats";
RL Gene 204:71-77(1997).
RN [3]
RP SEQUENCE OF 1-1132 FROM N.A.
RA Takeuchi T., Miesaki A.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: BINDS DNA NON-SPECIFICALLY.
CC -!- SUBUNIT: Part of the SWI-SNF complex.
CC -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -!- TISSUE SPECIFICITY: Highly expressed in spleen, thymus, prostate,
CC testis, ovary, small intestine, colon, and pbl, and at a much
CC lower level in heart, brain, placenta, lung, liver, skeletal
CC muscle, kidney, and pancreas.
CC -!- SIMILARITY: Contains 1 ARID domain.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1132.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF265208; AAF75765.1; ALT_INIT.

DR EMBL; AB001895; BAA23269.1; ALT FRAME.
DR EMBL; AB024075; BAA83073.1; ALT SEQ.
DR EMBL; AB024059; BAA83073.1; JOINED.
DR EMBL; AB024060; BAA83073.1; JOINED.
DR EMBL; AB024061; BAA83073.1; JOINED.
DR EMBL; AB024062; BAA83073.1; JOINED.
DR EMBL; AB024063; BAA83073.1; JOINED.
DR EMBL; AB024064; BAA83073.1; JOINED.
DR EMBL; AB024065; BAA83073.1; JOINED.
DR EMBL; AB024066; BAA83073.1; JOINED.
DR EMBL; AB024067; BAA83073.1; JOINED.
DR EMBL; AB024069; BAA83073.1; JOINED.
DR EMBL; AB024068; BAA83073.1; JOINED.
DR EMBL; AB024070; BAA83073.1; JOINED.
DR EMBL; AB024071; BAA83073.1; JOINED.
DR EMBL; AB024072; BAA83073.1; JOINED.
DR EMBL; AB024073; BAA83073.1; JOINED.
DR EMBL; AB024074; BAA83073.1; JOINED.
DR GENE; HGNC:11110; SMARCF1.
DR MIN; 603024; -.
DR GO; GO:0005634; C:nucleus; NAS.
DR GO; GO:0003677; F:DNA binding; NAS.
DR InterPro; IPR001606; ARID.
DR InterPro; IPR008938; ARM.
DR Pfam; PF01388; ARID; 1.
DR SMART; SM00501; BRIGHT; 1.
DR DNA-binding; Nuclear protein.
FT DOMAIN 96 99 POLY-GLN.
FT DOMAIN 178 184 POLY-GLN.
FT DOMAIN 615 618 POLY-SER.
FT DOMAIN 631 740 ARID.
FT DOMAIN 944 1021 GLN-RICH.
FT MUTAGEN 690 690 W-A: PARTIAL LOSS OF DNA-BINDING
FT ACTIVITY. WHEN A.A. 713 ALSO MUTATED,
FT COMPLETE LOSS OF ACTIVITY.
FT MUTAGEN 713 713 Y-A: PARTIAL LOSS OF DNA-BINDING
FT ACTIVITY. WHEN A.A. 690 ALSO MUTATED,
FT COMPLETE LOSS OF ACTIVITY.
FT CONFLICT 27 27 D -> G (IN REF. 1).
FT CONFLICT 51 51 V -> M (IN REF. 1).
FT CONFLICT 349 349 Q -> S (IN REF. 2).
FT CONFLICT 368 368 G -> GG (IN REF. 3).
FT CONFLICT 374 374 P -> S (IN REF. 2).
FT CONFLICT 393 393 P -> L (IN REF. 2).
FT CONFLICT 489 502 MANMPPQVSGSMCP -> NGQATSGWVRDVS (IN REF.
FT 2).
FT CONFLICT 534 534 E -> G (IN REF. 2).
FT CONFLICT 924 924 P -> S (IN REF. 2).
FT CONFLICT 1016 1016 Q -> L (IN REF. 2).
FT CONFLICT 1033 1033 Q -> P (IN REF. 2).
SQ SEQUENCE 1902 AA; 205946 MW; 9B70A46F81062EAA CRC64;
Query Match 2.5%; Score 332.5; DB 1; Length 1902;
Best Local Similarity 19.4%; Pred. No. 0.00059;
Matches 356; Conservative 202; Mismatches 664; Indels 609; Gaps 91;
Qy 728 GNEVPRCEGSPATVNNNS-----SDTESIPSPHTEAAKDTGQNGPKPATLGADGPP 779
Db 40 GSQTPQ---RYPMTVQGRASQSMGLSYTQIPYGGQSGSYGQGGQTP---YVYQSQSPH 94
Qy 780 PGPTTPPRTSRAPTEPTPASEATGATPPAPPS-PSAPPV--VPKKEKEETAAAPP 836
Db 95 PQQQPP--YSQQPPSQTTHAQSQQPQQPQQSQPQQSQPQQSQPQQSQPQQSQPQQSQP 152
Qy 837 VESEGEQ-----XPPAAEELAVDTGKAEPPVKSECTEAEAGPAKGDAAEAATAGAL 891
Db 153 SQSSTTQQHPQSPQSPYQQQPPQAPSTLSQQAAYPQPSQSQSQSQSQSQSQSQSQSQ 212
Qy 892 KAERKEGCGSRATKSSGAPQSDSSATCSADVDVAEGG--DKNRLSPRPILLPTG 949
Db 213 RFPFPQLS-----QUSFGQASSAPMTSSKGGQEDMNLQSRPSSL-----256
Qy 950 DPRANASPKPLDLKQLKQRAAAIPIQVTKVHEPPREDAA-----TKP-----AP 996

Db 257 -PDLGSGS-----IDDLPMGTEGALSPGVSTSGISSQGEQSNPAQSPFSPHTSPHLPGLGIRG 311
QY 997 PAPPP---PQNLOPESDAPQPCSSPRGKSRPAPPA-----DKEAFABAQKL 1042
Db 312 PPSPPVGSFASVAQSRGSLSPAAPV-GNQMPPRPPSGQSDSIMHPMSQSSIAQDRGYM 370
QY 1043 PGDP--PCWTSGLP-FPVPPREVIKASPHA-----PDPSAFSVAAPGHPL-PLGLHDTAR 1093
Db 371 QRNPMQPOYSSPQPCGALSPRQSGGQIHTGMSYQOQNSMGYSYGQGGYGPQGGY----426
QY 1094 PVLPRPTT-SNPPPLISSAKHPSVLERQIGALS-QGMSVQLH-----VPYS-----EH 1140
Db 427 ---PROQNTNALP---NANYPSA--GMAGGINPMGAGGQMGQPGIPYGYLPPGRMSH 477
QY 1141 A---KAPVGPVTMGLPLMDPKKLAPFSGVKEQELSPRQAGPPPSLGVPTAQEASVLRG 1197
Db 478 ASGNRPYGNMANMP-----POVG-----SGMCPPPGGMNRKT-QETAVAMH 519
QY 1198 TALGSPGGSITKGIPTSTPVSIDSALTIRGSIHTGTPADVLKGTITRIIGEDSPSLDR 1257
Db 520 VAANSI-----QNEPPGYPNMNGGMMGTGPP-----YQGINSMAGMINP-----Q 561
QY 1258 GREDSPKGVHVIYEGKKHVLSE-----GGMSTQCSKEDGRSSGPPHETAAPK-----1308
Db 562 GPPYSM--GGTMANNISAGMAASPMMGLDVKLTPTAKNNKADGTPTKTESKKSST 619
QY 1309 -----RTYDM-----MEGRVGRALSSNIEGLMGRALPERHSPHL-----KEQH 1349
Db 620 TTNEKITDYLGGPERKQWDRYLAFTTEKAMGNTLPAVGRKPLDLYLYSVKBE--677
QY 1350 HIRGSIQO-----GIPRSYVEAQEDYLRREAKLLKREGTTP 1385
Db 678 --IGGLTQVKNKWKRELATNLVNTSSSAASSLKKQYIQLYAF-----ECKIERGEDPPP 732
QY 1386 -----PPPSRDLTEAYKTOALGPLKLPKPAHEGLVATVKAGRSIHIEP 1429
Db 733 DIFAAADSKSQPKTQPPSPA-----GSGSMQGPOTPQSTSSMAEGDLKPPTP 782
QY 1430 REELRHTPELPLAPPLKSGSITQGTPLKYDTGASTTGSKKHVRSLIGSPGRTPFPVHP 1489
Db 783 AS-----THSQIPLPLPGMSRNSVGIQAFNDGSDSTFOKN--SMTPNFGYO-PSMNT 834
QY 1490 LDVMADARALRACVEE-----SLKSRPGT-----ASSGSGSIARGAP---VIVPELKG- 1535
Db 835 SDMMG-----RMSYEPNKDPYSGMRKAPGSDPFMSGQGNCGMDPYSAAGPGLNV 888
QY 1536 ---PROSPLYTHEDHGAPFAGHLPRGSPVTMEPTPL--OEGSLSSKASQDRKLTSTPRE 1591
Db 889 AMGPRQ-----HYPYGGPYDRVTRTFEGIGPEGNMSTG-APQNLMPSPNDS 933
QY 1592 TAKSP-----HST-----VPEHHPIHPITSP-----YHLLRGV 1618
Db 934 GWSYSPRYPPQQQQQQQRHDSYGNQFSTQCTPSGSPFPSSQITMYQQQQNYKRPMDGT 993
QY 1619 SGVDLYRSHIFLADPTSPIGRILPDAAAAYLPRHLAPNPTYPHLYPPVLIIRGYPTAA 1678
Db 994 YGPPAKRHEGEMYSVPYSTGQGPQO-----QQLPPAQOPQ-----ASQOQAAQ 1037
QY 1679 LENRQTIINDYITSQMHNTATAMAQRADMLRGLSPRESSALNYAAGRGIIDLQV 1738
Db 1038 PSPQDVYNOY---GNAYPATATATERRP-----AGGPO-----NQFP 1073
QY 1739 HLPVLVPTPGTATAMDRLAYLTPAPQPPFSRHSPLSPGGPHTLTKPTTSSSER 1798
Db 1074 F-----QFGRDRV-----SAPPGTNAQQNMPPQWGGP-----1101
QY 1799 DRDRERDREREKSLTSTTVEHAPITWRPGTEQSGSGSGSGSGSGSSRRPASHAH 1858
Db 1102 -----IQASAEVAQQTWQO-GRNDMTYNTYANRQSTGAPQGPAYHG---1142
QY 1859 QHSPISPRTOALQORPSVLHNTGMKGIITAVEPSKPTVLRSTSTSPVPAATFFPATH 1918

Db 1143 -----VNRTEMLHTDQANHEGWSHGTRQPPYGP-----SAPVPMTPRPPPSNY 1189
QY 1919 CPLGTTGLDVYPTILMEPVLLPKEAPRVARPERPRADTGHAFPAKPPARSGLPEASSPSK- 1977
Db 1190 QP-----PPSMQNHIPQVSSP-----APLPR-----MENRTSPSKS 1221
QY 1978 -----GSEPRPLVPVSGHATARTPAKNLAPHASP-----DPPAPPASASDPHREKTQ 2027
Db 1222 PFLHSGMMQKAGFPV-----PASHIAPAPVQPPMIRRDITFFPGSV-----BATQ 1267
QY 2028 SKPFSIOELELSLGRSYHSGSSYSPRGVFPVSPVSPSLTHD-----2067
Db 1268 PVLKORRELTKWDIG-----TPFAWRVMSLSKGLLAETWALDTINILYDDNSMTWF 1321
QY 2068 --KGLPKHLELD-----KSHLEGE-----LRP-----KQPGFVKLGG 2099
Db 1322 NLSQPLGLLELLVYFRCLIEIFGLKYEVDGPGQRTLLDPCGRFSKVSPPAMEGEE 1381
QY 2100 AAHL--PHL-----RPLPSQSPSSPLLOTAPOV-----KGQHV 2132
Db 1382 EELLGPKLEBEEVEEVEDEIAFGKDKPASENSEEKLISKFDKLPVKIVQKNDPFV 1441
QY 2133 VTLAQHISEV-----ITQDYTRHHPQOLSAPLPAPLYSPFGASCPLDLR 2177
Db 1442 VDCSDKLGRVQEPFDSGLLHWRIGGDTTEHIQTHFSKTELLP-----SRPHAPC- 1491
QY 2178 RPPSDLYLPPDHGAPARGSPHSEGGKRSRPNKTSVLGGEDGIEPVSPE-----GMT 2232
Db 1492 -PPA-----PRKVTTAEGTPTGT-----DQSGPPPDGPPKRRITATMD 1529
QY 2233 EPGHRSVAVPLLYRDQ-EOTEPSPRMGSKSP 2262
Db 1530 DMLSTRST---JTEDGAKSSEAIKESKPP 1557

RESULT 74

SON MOUSE

ID SON MOUSE STANDARD; PRT; 2404 AA.
AC OQX47; OQCO12; OQCK6; OQXPS;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE SON protein.
GN SON.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC STRAIN=129/Sv;
RX MEDLINE=20408886; PubMed=10950926;
RA Wynn S.L., Fisher R.A., Pagel C., Price M., Liu Q.Y., Khan I.M.,
Zammit P., Dadrah K., Mazrani W., Kessling A., Lee J.S., Bulwela L.;
RT "Organization and conservation of the GART/SON/DONSON locus in mouse
and human genomes."
RL Genomics 68:57-62(2000).
RN [2]
RP SEQUENCE OF 1-116 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Hippocampus, Small intestine, and Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Saito T., Okazaki Y., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Frieschmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincin S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

QY 1097 --PRPTTISNPPLISSAKHPSVLE--ROIGALSQMSVOLHVPYSEHAKAFVGPVTMGL 1152
Db 500 EQPVTTTEQPVAMTVVHPGHPEVTTATGLQG-----PEAAVVL 541
QY 1153 PLPMDPKLAPFSGVKQEQLSPRGQ---AGPPESLGVPTAQBASVLRGTALGS-----VPG 1205
Db 542 ELPGQB-----VATTALELGGQPSVTVGPVLSGLPSATRALELSGQSVATGALELPG 593
QY 1206 GSITKIPSTRVPSDAITYRSIHTGTPADVLYKGTTRIIGEDSPRLRGREDSLPK 1265
Db 594 QLMATG-----ALEFSQSGAAGALELLGQFLATGVL-----ELP- 628
QY 1266 GHVIVGKKGHVLSYEGGMSVTOCKEDGRSSGPPHETAAPKRTYDMMEGRVGRAISSA 1325
Db 629 -----GQPG-----APELPGQPVATVA-----LEISVQSVVTIS 657
QY 1326 STEGLMGRAPRPHRSHPHLKEQHIRGSIITGIPRSYVEAQEDYLRBREAKLLKREGTTP 1385
Db 658 ELSTM-----TVSQSLVSPSTTALESY-----NTVA 683
QY 1386 PPPSRDLTAVKTOALGPLKPAHEGLVATVKEAGRSIHIEPREEL-RHTPELPLAPR 1444
Db 684 QELPTLVGETSVTVGVDFLMAQEGHMLASNTMETHMLASNTMDSQMLASNTMDSQMLAS 743
QY 1445 PLKEGSITQGTPLKYDTGASTTGSKKHVRSVLIGSPGRTPFPVHPLDVMA---DARALE 1500
Db 744 NTMDSQMLASNTMDSQMLASNTMDSQMLASNTMDS-----QMLATSSMDSQMLA 792
QY 1501 RACYEESLKSPP-----GTASSGGGIARGAPVIVPELQKPSQPLTYEDHGAPFAGH 1553
Db 793 TSSMDSQMLASNTMDSQMLASNTMDSQMLASNTMDSQMLASNTMDSQMLASNTMDSQMLAS 840
QY 1554 LPRGSPVTWREPTPLRQSGSSSSKASQDKLTSPTREIAKSPHSTVPEHHPHISPYEH 1613
Db 841 LASGT-MAQASLTQDSAMGSKSPDPYRLAQDPYRLAQDPYRL-CHDPPYRLGHDAY 897
QY 1614 LLRGVSGVDLYR-SHIP--LAFDPSTI--PRGIPDLMAAAYLPRHLPNTPYHLYPYL 1669
Db 898 RL-----GQDPYRLGHDPYRLTPDPYRVSPR--PYRIAPRSY--RIAPRP----- 938
QY 1670 IRGYPDTAALNRQIINDITYSQM-----1695
Db 939 YRLAPRLMLASRRSMMSMYAERSMMSYERSMMSYERSMMSYERSMMSYERSMMSYERSMMS 998
QY 1696 --HNTATAMAORA-----DMLRGLSPRESSL 1720
Db 999 AYERSMMSYERSMMSYERSMMSYERSMMSYERSMMSYERSMMSYERSMMSYERSMMSYERSMMS 1058
QY 1721 ALNYAAGPRGII-----DLSQVPHLPVLVPPPTGTPATAM 1755
Db 1059 MSSYSAADRSMMSSYTDMSMAADSMTDSYTDSTYEAVMVPLPPEEPT----- 1110
QY 1756 DRLAYLPTAPOPFSSRHSSPLSGCPHLLTKPTTSSRERDRDRDRDREREKSL 1815
Db 1111 -----MP--PLPPEEPTPLPPEEPEPALST-----EQSAL 1143
QY 1816 TSTTVEHAPITWRFTEQSGSSGSSGSSGSSRASHAHQHSPISTQDALQORP 1875
Db 1144 TADNT-----WSTEVLTSTGESLS-----QPEPPVQSSEISESPMAV-PANTYSMSSET 1190
QY 1876 SVLHNTGKGIITAVEPSKPTVLSTSTSSVPRPAATFP-----PATCPLGLGLDGV 1928
Db 1191 SML-----ASEAVMTVPEAREP--ESSVTSAPVESAVVAHEWVPERPMTYVSETTMS-- 1243
QY 1929 YPTLMPEVLLKPEAPRVARPE-----RPRADTGHAF-----LAKP-----PARSGLE 1970
Db 1244 -----VEPAVLTSASVISETSETYDSMRP-----SGHALSEVTMSLLEPAVTSQPAEDSLE 1296
QY 1971 -----PASSPSKSGE-----PRPLVPPVSGHATIA--RTPAKNL-APHHA 2007
Db 1297 LPSMTVPAPSTMVTTTGPVAVTETBPPVAVPEPPIMAVPELPTMAVVKTPAVAVPEPLVA 1356

QY 2008 SPDP--APP-----ASADPHREKTQSPKPSIQEILERSLGYHGSYSVSPGV---EPVSP 2058
Db 1357 APPEPTTATPELCSLSVSEP-----PVAVSELPALADPEHAIT-AVSGVSSLEPSVP 1407
QY 2059 VSSSLTHDKLPHLELDELHLEGLRKPQGPVKGLEAAHLPHLRPLPESQPSSSP 2118
Db 1408 ILEFAV-----SVLQ-----PVMIVSEPS-----VPVOEP----- 1432
QY 2119 LLQTAQVKGHQRVVTLAQHISEVITQDTRHHHPQOLSAPLPAPLYSPFGASCPLDLRR 2178
Db 1433 -----TVAVSEPAVIVSEHT-----QITSPEMAV-----ESSPVI----- 1462
QY 2179 PPSDLYLPDPHGAPARGSPHSEGGKSPENKTSVLGGCGDIEP-----VSPPEG 2230
Db 1463 -----VDSVMSQQIMKGMNLLGGDENLGPVGMQETLLHPGEE 1501
QY 2231 MTEFGHSRSVAVPILLY-RDGEQTEPSPRMGSKSPONTSQPPAFTSKLTESAMVK--SKK 2287
Db 1502 PRDGGHLKSDLYENEDRNADLTWNHSLIVKDAEHTVCAITVGPVGEASEEKLPISET 1561
QY 2288 QEINKKLNTNRPENPEYNIQPGTEIPNMPAITGTGL-MTYRSQAVOE--HASTNMGLEA 2344
Db 1562 KEI-----TELATCAAVSEADIGRSLSQLALELDTVGTSGKFEF 1601
QY 2345 IIRKALM--GKYDQWEEPPPLSANAFNPLNASASLPAAMPITAADGRSDHTL-TSPGGG 2401
Db 1602 VTASALISESKYD-----VEVSVTTQDTEHDMVITSPSGGS 1638
QY 2402 KAKVGRSPSSRKAKSPAGLASGRPPSVSSVHSEGCNRRRTPLTNRVMBDRPSSAGSTP 2461
Db 1639 EADIEG-----PLPAKDIHL-----LPSTNFWCKVDSDSL--- 1669
QY 2462 FPNYPLIMRLQAGVMASPP 2480
Db 1670 -----PIKESAQAVAVALSPP 1684

RESULT 75
NFM_BOVIN
ID_NFM_BOVIN STANDARD; PRT; 810 AA.
AC O77788;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M) (Fragment).
GN NEF3 OR NEFM OR NFM.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hill W.D., Zhang L., Balin B.J., Sprinkle T.J., Spicer K.,
RA Gearhart D.A.;
RT "The bovine neurofilament M subunit has a novel set of KSP repeats
normally restricted to NF-H";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Neurofilaments usually contain three intermediate
filament proteins: L, M, and H which are involved in the
maintenance of neuronal caliber.
CC -!- PTM: There are a number of repeats of the tripeptide K-S-P, NFM is
phosphorylated on a number of the serines in this motif. It is
thought that phosphorylation of NFM results in the formation of
interfilament cross bridges that are important in the maintenance
of axonal caliber (By similarity).
CC -!- PTM: Phosphorylation seems to play a major role in the functioning
of the larger neurofilament polypeptides (NF-M and NF-H), the
levels of phosphorylation being altered developmentally and
coincident with a change in the neurofilament function (By
similarity).
CC -!- SIMILARITY: Belongs to the intermediate filament family.

RN [7]
RP SEQUENCE OF 1227-1289 FROM N.A.
RX MEDLINE=86104139; PubMed=3002437;
RA Nunez A.M., Francomano C., Young M.F., Martin G.R., Yamada Y.;
RT "Isolation and partial characterization of genomic clones coding for
RT a human pro-alpha 1 (II) collagen chain and demonstration of
RL restriction fragment length polymorphism at the 3' end of the gene.";
RN Biochemistry 24:6343-6348(1985).
[8]
RP SEQUENCE OF 1176-1226 FROM N.A.
RX MEDLINE=84118798; PubMed=6320112;
RA Strom C.M., Upholt W.B.;
RT "Isolation and characterization of genomic clones corresponding to
RT the human type II procollagen gene.";
RL Nucleic Acids Res. 12:1025-1038(1984).
[9]
RP SEQUENCE OF 35-167 FROM N.A.
RX MEDLINE=89233138; PubMed=2714801;
RA Su M.W., Benson-Chanda V., Vissing H., Ramirez F.;
RT "Organization of the exons coding for pro alpha 1(II) collagen N-
RT propeptide confirms a distinct evolutionary history of this domain of
RT the fibrillar collagen genes.";
RL Genomics 4:438-441(1989).
[10]
RP REVIEW ON VARIANTS.
RX MEDLINE=91184577; PubMed=2010058;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in collagen genes: causes of rare and some common diseases
RT in humans.";
RL FASEB J. 5:2052-2060(1991).
[11]
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;
RT "Mutations in fibrillar collagens (types I, II, III, and XI), fibril-
RT associated collagen (type IX), and network-forming collagen (type X)
RT cause a spectrum of diseases of bone, cartilage, and blood vessels.";
RL Hum. Mutat. 9:300-315(1997).
[12]
RP VARIANT SER-1074.
RX MEDLINE=90036909; PubMed=2572591;
RA Vissing H., D'Alessio M., Lee B., Ramirez F., Godfrey M.,
RA Hollister D.W.;
RT "Glycine to serine substitution in the triple helical domain of pro-
RT alpha 1 (II) collagen results in a lethal perinatal form of short-
RT limbed dwarfism.";
RL J. Biol. Chem. 264:18265-18267(1989).
[13]
RP VARIANT SEDC 1095-GLY--TYR-1330 DEL.
RX MEDLINE=89266907; PubMed=2543071;
RA Lee B., Vissing H., Ramirez F., Rogers D., Rimoin D.;
RT "Identification of the molecular defect in a family with
RT spondyloepiphyseal dysplasia.";
RL Science 244:978-980(1989).
[14]
RP VARIANT OSTEOARTHRTIS CYS-650.
RX MEDLINE=90370826; PubMed=1975893;
RA Ala-Kokko L., Baldwin C.T., Moskowitz R.W., Prockop D.J.;
RT "Single base mutation in the type II procollagen gene (COL2A1) as a
RT cause of primary osteoarthritis associated with a mild
RT chondrodysplasia.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:6565-6568(1990).
[15]
RP VARIANT OI-IV VAL-717.
RX MEDLINE=91291136; PubMed=2064612;
RA Bateman J.F., Hannagan M., Chan D., Cole W.G.;
RT "Characterization of a type I collagen alpha 2(I) glycine-586 to
RT valine substitution in osteogenesis imperfecta type IV. Detection of
RT the mutation and prenatal diagnosis by a chemical cleavage method.";
RL Biochem. J. 276:765-770(1991).
[16]
RP VARIANT OSTEOARTHRTIS CYS-650.
RX MEDLINE=91086471; PubMed=1985108;

RA Eyre D.R., Weis M.A., Moskowitz R.W.;
RT "Cartilage expression of a type II collagen mutation in an inherited
RT form of osteoarthritis associated with a mild chondrodysplasia.";
RL J. Clin. Invest. 87:357-361(1991).
[17]
RP VARIANT HYPOCHONDROGENESIS GLU-984.
RX MEDLINE=93054548; PubMed=1429602;
RA Bogaert R., Tiller G.E., Wies M.A., Gruber H.E., Rimoin D.L.,
RA Cohn D.H., Eyre D.R.;
RT "An amino acid substitution (Gly853-->Glu) in the collagen alpha
RT 1(II) chain produces hypochondrogenesis.";
RL J. Biol. Chem. 267:22522-22526(1992).
[18]
RP VARIANT HYPOCHONDROGENESIS SER-705.
RX MEDLINE=92262484; PubMed=1374906;
RA Horton W.A., Machado M.A., Ellard J., Campbell D., Bartley J.,
RA Ramirez F., Vitale E., Lee B.;
RT "Characterization of a type II collagen gene (COL2A1) mutation
RT identified in cultured chondrocytes from human hypochondrogenesis.";
RL proc. Natl. Acad. Sci. U.S.A. 89:4583-4587(1992).
[19]
RP VARIANT WS-II ASP-198.
RX MEDLINE=93304428; PubMed=8317498;
RA Koivikko J., Ritvaniemi P., Haataja L., Kaeeriaeinen H.,
RA Kivirikko K.I., Prockop D.J., Ala-Kokko L.;
RT "Mutation in type II procollagen (COL2A1) that substitutes aspartate
RT for glycine alpha 1-67 and that causes cataracts and retinal
RT detachment: evidence for molecular heterogeneity in the Wagner
RT syndrome and the Stickler syndrome (arthro-ophthalmopathy).";
RL Am. J. Hum. Genet. 53:55-61(1993).
[20]
RP VARIANT SEMD CYS-840.
RX MEDLINE=93282819; PubMed=8507190;
RA Holderbaum D., Malemud C.J., Moskowitz R.W., Haqgi T.M.;
RT "Human cartilage from late stage familial osteoarthritis transcribes
RT type II collagen mRNA encoding a cysteine in position 519.";
RL Biochem. Biophys. Res. Commun. 192:1169-1174(1993).
[21]
RP VARIANT SEMD ARG-285.
RX MEDLINE=93252400; PubMed=8486375;
RA Vakkula M., Ritvaniemi P., Vuorio A.F., Kaitila I., Ala-Kokko L.,
RA Peltonen L.;
RT "A mutation in the amino-terminal end of the triple helix of type II
RT collagen causing severe osteochondrodysplasia.";
RL Genomics 16:282-285(1993).
[23]
RP VARIANT SEDC CYS-206.
RX MEDLINE=94063862; PubMed=8244341;
RA Williams C.J., Considine E.L., Knowlton R.G., Reginato A., Neumann G.,
RA Harrison D., Buxton P., Jimenez S.A., Prockop D.J.;
RT "Spondyloepiphyseal dysplasia and precocious osteoarthritis in a
RT family with an Arg75-->Cys mutation in the procollagen type II gene
RT (COL2A1).";
RL Hum. Genet. 92:499-505(1993).
[24]
RP VARIANT SEDC CYS-920.
RX MEDLINE=93315508; PubMed=8325895;
RA Chan D., Taylor T.K.F., Cole W.G.;
RT "Characterization of an arginine 789 to cysteine substitution in
RT alpha 1 (II) collagen chains of a patient with spondyloepiphyseal
RT dysplasia.";
RL J. Biol. Chem. 268:15238-15245(1993).
[25]
RP VARIANT SEDC SER-1128.
RX MEDLINE=93140139; PubMed=8423604;
RA Cole W.G., Hall R.K., Rogers J.G.;


```
Db 491 STTESSAPVTSSSTESSAPVP-----TPSSSTESSA-PAPTPSSSTESSA 540
QY 1012 P-----QPGSSPRKSRSPAPPADKEAFAPAAQKLPDGPFCWTSGLPFPVP 1058
Db 541 PVTSSSTESSAPVPTPSSSTSSPTVTSSTESSAPV-PTPSSSTESSAPVPTP 599
QY 1059 PREVIK-ASPHADPDGAFSVAPGHPPLPLGLDHTARVLP-PTISNPPPLISSA 1112
Db 600 SSSSTESSAPATPSSSTESSAPVTSSTESSAPVPTPSSSTESSAPVPTPSS- 658
QY 1113 KHSVLERQTAISQCMVQLHVPYSEHAKAPVGPVTMGLPLPMDPKLAPSGVKQEOL 1172
Db 659 -----STTESSAPVPTPSSSTESSAPVTS-----TISSAPVTSSTESS 703
QY 1173 S-----PRGAGPPELSGVTAGASVLRGTALGVPGGSIYKIPSTRVPSDSAITVRG 1227
Db 704 SAPVPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTTE---SSAPVTSSTESS 760
QY 1228 SITHGTPADVLYKGTITRIIGEDSPSRLDRGREDSLPKGHVIEGKGHVLSEYEGMSVT 1287
Db 761 SAPVPTPSS-----STT 772
QY 1288 QCKED-GRSSGPPHETAPKRTYDMMEGRVGRAISSASIBGLMGRAPPERHSPHLK 1346
Db 773 ESSAPVPTPSSSTESSAPVPT-----PSSSTESSAPVPTPSSS----- 816
QY 1347 EQHHRIGSTQGIPTRS--VEAQEDVLRREALKREGTPPPPPPSRLTAYKTQALGP 1404
Db 817 -----NITSAPOSTPSSSTE-----SSSVPTPSSSTES-----SSAP 853
QY 1405 LKLKPAHEGLVATVKEAGRSIHEIPREELRHTPELPAPRLKEGSIQTGTPLYKDYTGAS 1464
Db 854 VS-SSTESSAPV-----PTSSSNIITSSAPSIPTP-SS 887
QY 1465 TTGSKKHVRSILGRCRTPFPVPHLDVMDARALACRYEESKSRGTPASSGGSTAR 1524
Db 888 TTES-----FSTGTITVP-----SSSKYPSQTETSVS-ST 917
QY 1525 GAPVIVP-----ELGKPRQSPITYE--DHGAPFAGHLPRG-SPVTMRPTPRLQEGSLS 1576
Db 918 TETTIVPTKTTSVTPTSTTITTVCTGTSNAGETSGSKPKVTITVTTTTSVT 977
QY 1577 SKASQDKLTSTPREIAKSPHSTVEHHHPISPVEYHLLRGVGVLYRSHIPLAFDPTS 1636
Db 978 SSTT---TITTT---VCSTGTSNAGE-----TTSGC-----SEPT 1006
QY 1637 IPRGIPDAAAYLPRHLAPNPTVPHLYPPYLINGYDPTAALENRQTIINDIYITSQMH 1696
Db 1007 ITTVCTSPSTSEASESTTSPTP-----VTVVSTTVTTEYSTSTKPG 1053
QY 1697 HNTATAMAQADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHLPVLVPPPTGPATAMD 1756
Db 1054 GEITTTFTVK-----NIPTTY-----LTTIAPTPTSVTNTFTPTTIT 1092
QY 1757 RLALYPTAPQFSSHHSSPLSPGCPHTLTKPTTSSSERDRDRDRDREREKSILT 1816
Db 1093 TVCSTGTN-----SAGETSGSKPKVT-TTVPCSTGTGTYTE-----ATLVT 1136
QY 1817 S---TTTVEHAPIWRPGTQSSG--SSGSGGGGSSSRPASH-SHAHQSPISPRDOTAL 1871
Db 1137 TAVITTV-----VTESSTGTSNAGKTTGYTKSVPTTYVTVTLAPAPVPTANAV- 1188
QY 1872 QORPSVLN-----TGMKGIIITAVEPSKPTVLRLSTSTSPVRPAATFPFATHCPLGGTL 1925
Db 1189 ---PTTITTTCSAATNAAGETTSVCSAK-TIVSSASAGENTAPSAITPTVTAIP----- 1239
QY 1926 DGVTPTLMEVLLKPEARVPRPRADTCHAFKAPPAR-SCLEPASPSKSGSEP-RP 1983
Db 1240 -----TTVITTES-SVGTNAGETTTGYTKSIPTTYITLIPGSGAKNYETVAT 1289
QY 1984 LVPPVSGHAT--IART-PAKNLAPHASPPPPAPPASAS 2019
```

```
Db 1290 ATNPISIKTSQLATTASASSVAPVVTSPSLTGPLQAS 1328
RESULT 79
IE18 PRIVIF
ID IE18 PRIVIF STANDARD; PRT; 1461 AA.
AC PL1675;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE Immediate-early protein IE180.
GN IE.
OS Pseudorabies virus (strain Indiana-Punkhauser / Becker) (PRV).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicellovirus.
OX NCBI_TaxID=31523;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=89315207; PubMed=2546124;
RA Cheung A.K.;
RT "DNA nucleotide sequence analysis of the immediate-early gene of
pseudorabies virus."
RL Nucleic Acids Res. 17:4637-4646 (1989).
RN (2)
RP REVISIONS.
RA Cheung A.K.;
RL Submitted (NOV-1989) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Nucleus of infected cells.
CC -!- PTM: A long stretch of serine residues may be a major site of
CC phosphorylation.
CC -!- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X15120; CAA33214.1; --
DR PUR; S04713; EDSEIF.
DR InterPro; IPR005205; Herpes_ICP4_C.
DR InterPro; IPR005206; Herpes_ICP4_N.
DR Pfam; PF03585; Herpes_ICP4_C; 1.
DR Pfam; PF03584; Herpes_ICP4_N; 1.
DR Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 390 405 POLY-SER.
FT DOMAIN 958 966 POLY-SER.
SQ SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 2.5%; Score 330; DB 1; Length 1461;
Best Local Similarity 21.0%; Pred. No. 0.00056;
Matches 366; Conservative 128; Mismatches 602; Indels 644; Gaps 86;

QY 673 MEKERNARKKKKAPAAASEEA-AFPVVVEDEMEASGVSGNE-----EEMVEEAE 722
Db 9 IETEGNFSQLAAAAAAFEEGIASGPDGSGSRRRSGSGEDLLFGPGLFSDDAEAE 68
QY 723 A---LHASANEVP-----RGECSGPAVNNSSDTEIPSPHTEAKDTGQ--- 764
Db 69 AAVLAAAGATPPPPPSAQOORHARRG--SGEIVLDDDEEE--DEPGSPAAGSFGAL 125
QY 765 -----NGPKPP-----ATLCADGPP----- 779
Db 126 HQSGEHLVLPGRSAGSGPRPTTAAALAAEACAPGPGRSPSSPASPASSSSGSGS 185
QY 780 ---PGPPTTPRTSA---PI-EPTPASEATGPTTTPPAPSPSAPP-----PVVPKEKE 828
```


RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S.C., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Grahm D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.B., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leivaeslahti M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., McLeay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkneen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A.C., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE OF 105-1083 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=97098442; PubMed=8942982;
RX Tanese N., Saluja D., Vassallo M.F., Chen J.-L., Admon A.;
RT "Molecular cloning and analysis of two subunits of the human TFIIID
complex: hTAFII130 and hTAFII100.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13611-13616(1996).
RN [4]
RP IDENTIFICATION IN THE TFIIIC-HAT COMPLEX WITH TAF5L; TAF6L; TADA3L;
RP SUPT3H; TAF2; TAF5; TRAP; GCN5L2 AND TAF10.
RX MEDLINE=9303588; PubMed=10373431;
RX Brand M., Yamamoto K., Staub A., Tora L.;
RT "Identification of TATA-binding protein-free TFIIIC-containing complex
subunits suggests a role in nucleosome acetylation and signal
transduction.";
RL J. Biol. Chem. 274:18285-18289(1999).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 870-918 IN COMPLEX WITH
RP TAF12.
RX MEDLINE=20063193; PubMed=10594036;
RX Gaglioffi Y.-G., Werten S., Romier C., Carre L., Poch O., Moras D.,
RA Davidson I.;
RT "The human TFIIID components TAF(II)135 and TAF(II)120 and the yeast
RT SAGA components ADA1 and TAF(II)68 heterodimerize to form histone-like
RT pairs.";
RL Mol. Cell. Biol. 20:340-351(2000).
CC -!- FUNCTION: Makes part of TFIIID is a multimeric protein complex that
CC plays a central role in mediating promoter responses to various
CC activators and repressors. Potentiates transcriptional activation
CC by the AP-2 of the retinoic acid, vitamin D3 and thyroid hormone.
CC -!- SUBUNIT: TFIIID is composed of TATA binding protein (TBP) and a
CC number of TBP-associated factors (TAFs). Component of the TFIIIC-HAT
CC complex, at least composed of TAF5L, TAF6L, TADA3L, SUPT3H/SPT3,
CC TAF2/TAFII150, TAF4/TAFII135, TAF5/TAFII100, GCN5L2/GCN5, TAF10
CC and TRAP.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- SIMILARITY: BELONGS TO THE TAF2C FAMILY.
CC -!- SIMILARITY: Contains 1 TAFH/NHR1 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y11354; CAA72189.1; -;
CC EMBL; AL137077; CAC36006.1; -;
CC EMBL; AL109911; CAC22312.2; -;

DR EMBL; U75308; AAC50901.1; -;
DR PDB; 1H30; 26-SEP-02.
DR TRANSFAC; T02328; -;
DR Genew; HGNC:11537; TAF4.
DR MIM; 601796; -;
DR GO; GO:0005669; C:transcription factor TFIID complex; TAS.
DR GO; GO:0016251; F:general RNA polymerase II transcription fac. .; TAS.
DR GO; GO:0005515; F:protein binding; TAS.
DR GO; GO:0003713; F:transcription co-activator activity; TAS.
DR InterPro; IPR007900; TAF4.
DR InterPro; IPR003894; TAF_hom.
DR Pfam; PF05236; TAF4; 1.
DR SMART; SM00549; TAFH; 1.
KW Transcription regulation; Nuclear protein; 3D-structure.
FT DOMAIN 590 682 TAFH/NHR1.
FT DOMAIN 39 42 POLY-HIS.
FT DOMAIN 52 57 POLY-ALA.
FT DOMAIN 98 101 POLY-GLY.
FT DOMAIN 142 148 POLY-ALA.
FT DOMAIN 268 275 POLY-PRO.
FT DOMAIN 331 337 POLY-ALA.
FT DOMAIN 680 683 POLY-PRO.
FT DOMAIN 808 813 POLY-ALA.
FT DOMAIN 828 831 POLY-ASP.
FT CONFLICT 105 117 PGPPRRRLVPA -> GRGLLQORGGRES
FT CONFLICT 136 136 A -> S (IN REF. 2).
FT CONFLICT 185 185 G -> GPG (IN REF. 2).
FT CONFLICT 233 264 MISSING (IN REF. 3).
FT CONFLICT 293 293 P -> L (IN REF. 3).
SQ SEQUENCE 1083 AA; 109943 MW; A6453827572A0752 CRC64;
Query Match 2.5%; Score 329.5; DB 1; Length 1083;
Best Local Similarity 23.08; Pred. No. 0.00044;
Matches 230; Conservative 94; Mismatches 336; Indels 341; Gaps 50;
Qy 701 EDEMEASGVSGNEEEMVEEAEALH-----ASGNEVPRGSCSGPATVNN 744
Db 16 EVDEKVVSLVGSLSQLAASAAHHHLA PRTPEVRAAAAGALGNHVVSVPAGAAGAGP 75
Qy 745 SSOTESTPSHTEAAKDTGQNGKPP-----ATLGADGPP-PGPPTPPRTTSRAPIETPA 799
Db 76 AAPAAGAPGA-----APEPPAGRAPGPGGPGPGPGKPGAGGAAQTLLNGSAAL 117
Qy 800 SEATGAPTPPPPPSAPPVVPKEEKEETAAPVVEEGEEQKPPAAEELAVDTGKAE 859
Db 118 -----GPAPPAKLRRP-----PEGSAGACAPV-----PAAAAVAAG----- 149
Qy 860 EPVKSECTEAEAGPAK--GKDAEAEAA-----TAEGALKAEKKEGGSGRATTAKSGAP 912
Db 150 -----PEPAPAGPAGPAGPAAALAAAGPAGPGPGPGPGKPGAGGAAQTLLNGSAAL 202
Qy 913 QDSSTATCSADEVDEAGGDKNLLSPRS-----LLTP-----TGDPRANASQKPLDL 963
Db 203 LNSHAA---APAVSLVNGGPAALLPLPKPAAPGTVIQTTPFVGAAPPAAPSPPAAP 259
Qy 964 KQLQRAAAITPIQV--TKVHEPPREDAAPTKAPAPPPPPQNLQPEDAPQPGSSPRGK 1022
Db 260 AAPAAPAAPPPPPAPATLARPCHGAPPTTAAVAPPPA-----AAQNGSA--GA 309
Qy 1023 SRSPAPPADAEFAAEAKLFGDPFCWTSGLPFFVP-----PREVIKASPHA----- 1069
Db 310 APAPAPAGGPA-CVSGOPGPG-----AAAAAPAGVKAESPKRVQAPPAQTLLAAS 362
Qy 1070 -----PDPSAFSVAPPCHPLPL-----GLH 1089
Db 363 GPASTAASWVIGPTMQGALFSPAAPPAPGPTGLPKGAAGAVTQSLSRTPPTATTSGIR 422
Qy 1090 DTAEPV-----LPRPPTISNP-----PP--LISSAKHPSVLEQIGALSQGMVOLH 1134
Db 423 ATLTPVTVALRPLQPP--QNPTNTQNFQPLPGMVLVRSNGQLMLIQQALAAQ-MQAQAH 479
Qy 1135 V-PYSEHAKAPVGPV-----TMGLP-LPMDPKKLAPFSGVKOE-----Q 1171

[illegible]

RESULT 82

MD6C HUMAN STANDARD; PRT; 1003 AA.
ID MD6C HUMAN Q96DN6; Q9N3M0; Q9NA81; Q96Q00;
AC Q96DN6; Q9N3M0; Q9NA81; Q96Q00; Created
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Methyl-CpG binding domain protein 6.
GN MD6C OR KIAA1887.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RC TISSUE=Brain;
RA Takahiro H., Yamazaki M., Watanabe K., Kumagai A., Itakura S.,
RA Fukushima Y., Fujimori Y., Komiyama M., Sugiyama T., Irie R.,
RA Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA Kawai-Hiro Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Wagatsuma M.,
RA Murakawa K., Kanehori K., Takahashi-Fujii A., Ohnuma A., Sugiyama A.,
RA Kawakami B., Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K.,
RA Isogai T.;
RT "NEDO human cDNA sequencing project."
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21456161; PubMed=11572484;
RA Nagase T., Kikuno R., Ohara O.;
RA "Prediction of the coding sequences of unidentified human genes. XXI.
RT The complete sequences of 60 new cDNA clones from brain which code for
RT large proteins.";
RL DNA Res. 8:179-187(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Amygdala;
RA Wambutt R., Heubner D., Mewes H.-W., Weil B., Wiemann S.;
RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 1 methyl-binding (MBD) domain.

Query Match 2.5%; Score 325; DB 1: Length 1781;
Best Local Similarity 18.7%; Pred. No. 0.00096;
Matches 417; Conservative 273; Mismatches 727; Indels 810; Gaps 102;

QY 75 GNRSELHLRPSHLYLPELG-----KSEMEFI-----ESKRPRLELLPDPILL 118
DB 138 GQENRNIEQIPSSSENLBELTQPTBSQANDIGFKVKFKVGFVKTKDKTE-KPDIVQ 196
QY 119 RPSPLATGQAGSGLTWDRLTGLKLEVPSPPHDPELELVPRLSKELTONDR 178
DB 197 LUTVKDEGEGAGADHQPDL-GAGEAASESEPKOSTE-----KPEETLKREQ 246
QY 179 VDREITMVQOISKLKKQQLLEEAAPPEPEKVPSPD-----PIESKRSRLVLIIDEN-- 234
DB 247 SHAELSPAESGAVEEKEEKEEKEE--EFSKSAESPTSVTSETGSTKFTTQGA 304
QY 235 ---RKAABAAHRIEGLGPOVELPLYNQPSDTRQYHENTKINOAMRKLILYFKRNHARK 292
DB 305 GWRKKTST-----FRPKKEDV-----EASEKK-----KE 328
QY 293 QWKQPCORYDQLEALEKKVERIE-----NNPRRAKSKVREYVEKOPPEIRKORELOE 348
DB 329 QPEKVDTEDEKAVASEKLTAHQAPQEPESAHEPRISAEBYK--VELPSEEQVS- 385
QY 349 RMOQSVQOQSGLSAARSEHEVSEIIDGLSE-----QENLEKQWOLAVIPP 397
DB 386 -----GSQCPSEKPAPLATEVFEDEKIEVHQEEVVAHVSTVEETEEQTEV----- 434
QY 398 MLYDADOQRIKFINNGLMADPMKYKORQVNMWMSBOEKEFP-----REK 443
DB 435 -----ETAGSVPAELVGMDAEPQEAEPKLVK-----KETCVSGEDPTQCADLSPDEK 486
QY 444 FMQHPKNFGLIA-----SFLERKTVAECVLYLYLTKKNYKSLVRSYRERGSKSQOQQ 498
DB 487 VLSKPE-GVYSEVEMLSQEMKVGQSPKLKLTSTGLKLSKKQKGGKGGDEESGE 545
QY 499 QOQOQOQOQOQOQMPRPSOEKDEKEKEKEEKEPEVENDKEDLLKEKTDTSGEDN 558
DB 546 HTQ-----VPADSPDQSEQKQESSASSPEEPETICLEGLAEVQ-----QDG 589
QY 559 DEKAVASKRKTANSQCRKRG-----RITSMANEANSEEAITPOQSAE 603
DB 590 EAEGATSDGEK-----KRGVTPWASFKNWTPKKRVRR--PSESDKEDLDKVKSAT 641
QY 604 LASMELNESSRWTEEMETAKGLLEHGR-----NWSAIAARMVGSKTVSQCKN 651
DB 642 LSSTES-----TASEMQEEMKGSVEEPKPEPKRVDTSVSWEALI-CVGSS----- 687
QY 652 FYFNVKBQNLDEILQHKLWKEKERNARRKKKAPAAASEEAAPPPVV----- 700
DB 688 -----KRR-----ARR-----SSDEEGGPKMGGDHOKADEAG 717
QY 701 EDEMEASGVGBEEMVEEAALHASNEVPRGECGPATVNNSSDT-ESITSPHTEAA 759
DB 718 KDKETGTGILAGQE-----HDPQSGSSPEQAGSPTGEGVSTWSEFKRLVTPRK 769
QY 760 KDTQNGKPKPATLGAGDPPPPPP-----TPPRRTSR-----APIE- 795
DB 770 KSKSLEKESDSIAGSGVEHSTPDTEPGKEESWVSITKFTPGRRKKRPGDKQEQAPVED 829
QY 796 --PTPASEATGAPTPPAPPSAPPVPVPEE---KEEETAAPPVVEGEEO-KPPAAE 849
DB 830 AGPTGANE-----DDSDVPAVPLSEYDAVEREMEAQQAQKGAPEQKAAAT 877
QY 850 ELAVDTGKAE-----EPVKSEC-----TEE--- 869
DB 878 EVSKELSESQVHMMAAVALDGTAAATIIERSPSWISASVTEPLEQVEAEALLTEVLE 937
QY 870 -----AEEGP-----AKGDAEAAEAAGALKAEKKEGSGRATTAKSSGAPQSDSS 918
DB 938 REVIAEEPTVTPELPENREARGDVTVSEALTEPE-----AVTAAETAGPLGS--- 986
QY 919 ATCSADEVDEAEGDKNRLSPRLTPTGDPANASPOKPL-----DLKQLKQRAAAI 973

DB 987 -----BEGTEASAEETEMVSAVSQLTSDPDTTEATPVQEVGGVPDIEEQERRTOEV 1041
QY 974 PPIQVTKVHEPPREDAAPTKPAAPPQNLQOPESDAPQOPGSSPRGKSRAPPADKE 1033
DB 1042 LOAVAELKVBESQ-----LPCTGGPEDVLQ-----PVORAEAPERPEQAESAGLUKE 1088
QY 1034 ---AFAABAQKLPQDPPCMTSGLPFPVPPPREVIKASPHADPPSAFSAVPPHPLPLGLHD 1090
DB 1089 TDVVLKVDAAQEAKEP-FTQG-----KVVQGT-----TPESFEKAP----- 1123
QY 1091 TARVPLRPPPTISNPPPLISSAKHPSVLEROIGAISQGMVQLHVHPVSEHAKAPVGPVTM 1150
DB 1124 -----QVTESIESSE----- 1133
QY 1151 GLPLPMDPKLAPSGVQKQSPRGQAGPPESLGVPTAQEASVLRGTALGSGVPGGSITK 1210
DB 1134 -LVTTQAEULA--GVKSQEMVME-QAIPDSEVETPDTSETDGTGTFVADFDAPGTTQKD 1188
QY 1211 GIPSTRVPSDAITYRGSITHGTTPADVLYKGTITRIIGEDSPSRLDRGREDSLPKGHVY 1270
DB 1189 EIVEIHENEVA--SGTQSGGTEAEAV-----PAKER---PPAPSFVFEQ 1229
QY 1271 EGKK-----GHVLSYEGGMSVTQCKEDGSSSGPPHETAAPKRYDMMEGVGRA 1321
DB 1230 EETKQSKMEDTLEHTDKEVSVETVILSKTEGTQEA---DOYADEKTKOV----- 1277
QY 1322 ISSASIEGLMGRAIPPERHSPHLKEQHHRISITQICIPRSYVQAEQEDVLRRE----- 1374
DB 1278 ---PFFEGL-----EGSIDGTIVTSREKVTVEALKGEGTEAEAC 1313
QY 1375 ---AKLLKREGTPPPPPSRDL---TEAYKTOALGPLKLKPAHEGLVATVKBAGRSIHE 1427
DB 1314 KDDALELOSHAKSPSPVEREMVQVREKTEA-----EPH----- 1351
QY 1428 IPREELRHTEPLAPRLKEGSITQCTPLKYDTGTASTTGSKKHVDVRSILIGSPGRTFPV 1487
DB 1352 VNEEKLEHETAVTVSEVSKOLLQTVNVI-----IDGAK--EVSSLEGSP---PPC 1398
QY 1488 HPLDVMDARALERACYEESLSKSRPG---TASSGGSIGARGAPVIVPELGPROSPLTY 1543
DB 1399 LGQE-----EAVCTKIQVOSSEASFTLTAABEEKVL-GETANILETGE-----TL 1443
QY 1544 EDHGAPFAGHLDRGSPVTWRETPRLQEGSLSSSKASQDRKLTSTPRETAKSPHSTVPH 1603
DB 1444 E-----PAGAHVLVEE-----KSEKNED--FAAHPGEDA-----VP-- 1473
QY 1604 HPHIPSPYBHLRGVSGVDLYRSHIPLAFDPTSIIPRGIPLDAAYVYLPRLHAPNPTVPH 1663
DB 1474 -----TGPDCAKSTPVIVSATT-KKGLSSD----- 1498
QY 1664 LYPPYLIRGYPDTALENRQTIINDYITSOQMHNTATAMAQRADMLRGLSPRESSLAIN 1723
DB 1499 -----LEG-EKXTSLKWKSDVDEQVACQEVKVSVAIEDLEPENGIILETKSKLVQN 1551
QY 1724 YAAGPRGIIDLQVPHLPVLVPTPGTATAMDRLAYL-PTAPOPFSSRHSSSPLSPGPG 1782
DB 1552 -----IIQ-----TAVDQVTRTEETATEMLTSELQTO----- 1578
QY 1783 THLTPTTSSSERDRDRDRDREREKSLTSTTTTTHAPIWRPGTEOSSGSSGSSG 1842
DB 1579 AHVIRKADS-----QDAGQETEKEGEBPOA-----SAQDETPIITSAKESESTAVGQA- 1625
QY 1843 GGGGSSSRPASHAHQHSPISPRTQDALQORPSPVLHNTGMKIITATPEPSKPTVLRSTS 1902
DB 1626 -----HSDISKOMSEASE-----KTMVTEVE----- 1646
QY 1903 TSSPVRPAATPPATHCPGLGTGCVPTLMPEVLLPKE-----APRVARPERPRADTGH 1957
DB 1647 -----GST---VNDQQLVEVLPSEEGGAGTKSVPE---DDGH 1680
QY 1958 AFLAKPPARSGLPASPSPKSGSEPRPLVPVPSGHATARTAKNLAPHASPDPPAPPAS 2017

Db 1681 ALLAEIRTEKSLVPEP-KEDEKGD---VDDPENQNSALADTDASG-GLTKESPDTNGP--- 1732

Qy 2018 ASDPHREKTSKPSIOEILRSILGYHSSYSPGVEFVSPVSSPSLTHDKGL-PKHLEE 2076

Db 1733 ---KQEKEDA---QVELOQEKVHSES-----DKAITPQAOEE 1765

Qy 2077 LDKSHLE 2083

Db 1766 LQOQERE 1772

RESULT 84

TEGU_EBV

ID TEGU_EBV STANDARD; PRT; 3149 AA.

AC P03186;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 01-APR-1993 (Rel. 25, Last annotation update)

DE Large tegument protein.

GN BFLF1.

OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Lymphocryptovirus.

OX NCBI_TaxID=10377;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84270667; PubMed=6087149;

RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,

RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,

RA Tuffnell P.S., Barrell B.G.;

RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";

RL Nature 310:207-211(1984).

CC -!- FUNCTION: Tegument protein.

CC -!- SIMILARITY: BELONGS TO FAMILY THAT GROUPS TOGETHER HSV-1 UL36,

CC EHv-1 24, EBV BFLF1, HVS-1 64, VZV 22, AND HCMV UL48.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>

CC or send an email to license@isb-sib.ch).

CC -----

CC EMBL; V01555; CAA24839.1; -.

DR PIR; G93065; Q0BE8.

DR InterPro; IPR006928; Herpes_teg_N.

DR Pfam; PF04843; Herpes_teg_N; 1.

SQ SEQUENCE 3149 AA; 337954 MW; 3DD0C576587313D8 CRC64;

Query Match 2.5%; Score 324; DB 1; Length 3149;

Best Local Similarity 19.2%; Pred. No. 0.0018;

Matches 510; Conservative 261; Mismatches 864; Indels 1026; Gaps 120;

Qy 92 LPEL-----CKSEMEFTESKRPLELLPDLPLRSPILLATGPAGSEDLTKDRSLTGKLE 146

Db 1098 LPSEFVAKQKELETTRENEKRLRTLLDDIEMGLAGVASAPGA----- 1143

Qy 147 PVSPSPPP-----HTDPE-----LELVPRLSK-----EELI 173

Db 1144 PISPAFSPATPANHDNPEATPTPLADTAALTPIVIEKIVIANAGSIVGAKNPTVIRLDDTI 1203

Qy 174 QNMDR-----VDREITMVQOISKLKKQOQLEEAAKPEPEKPVSP 217

Db 1204 QQIVRSKYLMLNLKSTFTYDNYIASPESIDHLYRDLPLVDLPEV--QDGDRLDPM 1261

Qy 218 PIESKUR-----SLVQIYYDENRKAEEAHRILEGL-----GPQ 251

Db 1262 VSEALHTFMGNRLTLEPARLVALQNPATHSLTKETAAAVNLLPGLLAVYDATITGQAE 1321

Qy 252 VEL-----PLYNQSDTRQYHENTKINQAMRKILILYF-KERNHARQWKQKFCQYDQLM 306

Db 1322 DALRLLSGLQNQLSQT-----LIPGKLKRFSLYLQKLKNNNDQLRQKVEQAM----- 1370

Qy 307 EALEKKVERIENPRRAKESKVREYYEKOFFEIRKQELQERMQSRVQSGLSMSAA 366

Db 1371 -----RLEAGFKPATEEQ-----EAFLDTPNKKELKQVEKKLRQ-----LMTG 1412

Qy 367 RSEHVESEIIDGLSEQENLEKQMRQLAVIPMLYDADQOQRIKFINMGLMADPMKYVKDR 426

Db 1413 RKEKE-----KLREQEDKERERRAR-----EANEAWARIRKALGARPEAPTSPDD 1459

Qy 427 QVMNMSEQSKETFREKFMQHPKNFGLIASFLERKTVAECLVYYLTKQENYKSL----- 482

Db 1460 -----WNT-----LIASLLPDNTDSAAAAAARNTDILDSLTQIL 1496

Qy 483 -----VRSYRRRGKSQQ 532

Db 1497 AAMLLGITRVRERLRS-----LLVDDGGAARMAEAPGWFDT 1535

Qy 533 -EKEPEVENDKEDLLKEKTDGTDGENDKEKAVASKOKRTANSQGRKGRITRSMANEAN 591

Db 1536 IETGPLARLDAMPATPAATAKEGGGGRGABEAGALFR-----ARTAADAIR 1582

Qy 592 SEEAITPQ--QSAELASMEINLESSRWTEEMETAKKGLLEHGRNWSAIAIRMVGSKTVSQC 649

Db 1583 SALAQTRQALQSPDMKSAVVNTDLAPYAEYERGLAGLEKRRABEALTAIVSEYVDR 1642

Qy 650 KNFYFNYKKRQNL--DEILQOHLKMEKERNARRKKKAPAAASEEAAFPVVE----- 701

Db 1643 LPEATNDPGQANLPPPTPIQ-----ATAPRLASDSALWPKPKQLLTRRE 1688

Qy 702 -DEMEASGVSGNE-----EMVEBAEALHAGNEVPRGECGSPATVNNSSD 747

Db 1689 RDDLQATGDFSELLTEAEAEVRALEEQVRESQTLMAKAHMAASTRRGFHTA----- 1743

Qy 748 TESISPSHTEAAKDTGQNGPKPPATLGADGPPGPPPTPRRTSRPTSEPTPASATCAPT 807

Db 1744 LEAVLSRSRDEAPD-----DELSLLPSPKAPVQAPLEAALARAAGNS 1789

Qy 808 PPPAPPPSPAPPPVPVPEKEEETAAAPVE--EGEEQKPPAAEEELAVDTG-----KA 858

Db 1790 WP-----YKSLAAAKWIRGICEAVRGLSEGALALAGGAGAWNLAAA 1832

Qy 859 EEPVSECTEAB-EGPAKGDABABATAGALKAEKKEGGGGRATTAKSSGAPQSDS 917

Db 1833 ADGEIHETLRLLEVEGMAQ-NSMDGMEELRLALATLDPKRVAGGKETVA-----DW 1882

Qy 918 SATCSADEVDEAGGDKNRLSPRSLTPT--GDPR-----ANASPOK 959

Db 1883 KRRLSLREAIQEAQESQLQGLTQDLVTOARGHTDPRQLKIIVVEAARGLALGASGSQY 1942

Qy 960 PLDLKQLKQRAAA-----IPPIQVTKVHEPPREDAAATPKAPPAPP-----PPQNL 1005

Db 1943 ALLKDKLLRVASAKQSLAFYETAQPTVFVKHPLTNLPLITISAPPTGNGCAPTRRA 2002

Qy 1006 QPESDA-POQPGSSPGKSRSPAPPADKAEFAABAQKLPGDPCWTSGLPVPPPREVIK 1064

Db 2003 QFLAAAGPAKYAGTLMLETESPCDPLNPAYVSADTQPLNYIPVYHNFLEYVMD--TVLE 2060

Qy 1065 ASPHAPDPSPAFSVPCHPLPLGLHDTARPVLPRPT----- 1101

Db 2061 -----NPEAFSLTPAGRPOAIG-----PPODQERRRRRTLASVASARLSAAA 2103

Qy 1102 -----ISNPPPLISSAKHPSVLERQIGAISSQMSVQLHVPVSEH 1140

Db 2104 DSYWDTWPDVSNAGELLREYVSAPKALMEDLADNPVAMTLLAHASLIASRNHPYP-- 2161

Qy 1141 AKAPVGPVT-----MGLPLMPDKKLAPFSGVKQEQSLSPRGQAGPPESLGVPT 1188

Db 2162 -----APATDREVILLEQREMMALLVGTHPAYAAAFGL-----APSFYAG--LGLVS 2206

Qy 1189 AQEASVLRGTALGVSFGSGSITKGIPTSTRVPSDSAITVR-----GSITHGTGPAVLVYKGTI 1243

Db 2207 A-----LARDGGLGD-----LLSDSVLTLYRLVRSPASGRGMPSTT--RGSN 2246

QY 1503 ----CYEESLSRFGTASS-----SGSARGAPVIVPELG-----KPRQSPLT----- 1542
Db 838 GSGPSSPASTKSSSTKSSSTKSGLSGSGYASSPAAGDPDPAPERKKRRAPGARRPG 897
QY 1543 --YEDHGAPAGHLPGSPVYMEPT-----PRLOEGL-----SSSKASQD 1582
Db 898 DGEDEG--LSGAALRGDGHGHDDEDRGPRKRRLSLGLGPAPDPAPALLSSSSSSD 955
QY 1583 RKLSTPREIAKSPHSVTPPEHHPIPIPYEHLRGVSGVDLYRSHIPLAFDPTSIPIRGIP 1642
Db 956 DRU-----RRPLGMPPE-HAPDGGFRVPA-----ETHTPR--P 988
QY 1643 LDAAAAYLP-----RHLAPNPTPHLYPPYLIIRGYPTAALENRQTIINDYITSQQMHN 1698
Db 989 SEAAALAYCPPEVARALVDQVPELWRPALTP-----DPAALA-----HIAARR----- 1033
QY 1699 TATAMAQADMLRGLSPRESLALNYAAGPRGIDLSQVPHLPVLVPTPTGTATAMD-R 1757
Db 1034 -GAPLRRAAMRQIADPED-----VRVVVLYDPLPHEELCAAPAEAGAPPAWDPR 1083
QY 1758 LAYLPTAQPFSSR-----HSSSPLSPGGPT--HLTKPTTSSERE----- 1797
Db 1084 RGLSALLAAFAHLRCLTPDGHAWAGNWTGRPDIGRLNAQVLLSLARDLGFAGAVEYLC 1143
QY 1798 -RDRDRDRDREREKSLTSTTTVEHAPITWRPGTEQSSGSGSGSGSGSGSGSRPASHSH 1856
Db 1144 RLGAARER-----LIVDTIEDWPADGPA-----VGDYH 1172
QY 1857 AHQSPISPTQDAL-----QORPSVHNTGKMG--IITAVEPS----- 1893
Db 1173 VYVRVLDPAQAQAVRWPGCRELRAAVLDSSIVGPAFCARVEASFARLHPGAEPRLCLR 1232
QY 1894 KPTVLARSTSSVPRPAATPPATHCPLGGLDGVVPTL--MEPVLLPKEAPRVARPERP 1951
Db 1233 QDNVRYTVSTRAGRTVPVLPFRAY-----RORVLTVDGCKDMARQSRALGLGDPD-- 1284
QY 1952 RADTGHAFIAKPPARSGLPASP-----SKGSEPRP-----LVPPVS 1989
Db 1285 -FDGAAGFGRHRAANRWGLGAPLPRFVSCGRRGLAELRPGGLPAELRAFCAALLEPDA 1343
QY 1990 GHATITARTPAKNLAPHASDPDPAPPASADPHREKTQSPFSIQLELESLG-----Y 2043
Db 1344 EAAPLVLTGC-----AVAAAGAPPAVLWD-----FAPFETSVRAAAGGAVETH 1386
QY 2044 HGSYSPEGVPEVPVSSPSLTHDKGLPKHLEELDKSHLEGLRKPQ-CPVKL 2096
Db 1387 RPAGASGAGAGPGEDGDSVEIVVRG-----GDGRPRGLGPIKV 1426

RESULT 87
MY33 HUMAN
ID MY33_HUMAN STANDARD; PRT; 2004 AA.
AC Q92794;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE MY33 histone acetyltransferase 3 (Runt-related transcription factor
DE binding protein 2) (Monocytic leukemia zinc finger protein) (zinc
DE finger protein 220).
GN MY33 OR RUNXBP2 OR ZNF220 OR MOZ.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96376968; PubMed=8782817;
RA Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA Chaganti R.S.K., Clavin C.I., Distche C., Dube I., Frischauf A.M.,
RA Horsman D., Mitelman F., Volinia S., Watmore A.E., Hausman D.S.;
RT "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT a putative acetyltransferase to the CREB-binding protein.";

RL Nat. Genet. 14:33-41(1996).
CC -I- FUNCTION: May represent a chromatin-associated acetyltransferase.
CC -I- SUBCELLULAR LOCATION: Nuclear.
CC -I- DISEASE: Involved in acute myeloid leukemias through a chromosomal
CC translocation t(8;16)(p11;p13) involving MYST3 and CREBBP.
CC -I- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC -I- SIMILARITY: Belongs to the MYST (SAS/MOZ) family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL; UA7742; AAC50662.1; --
CC Genew; HGNC:13013; MYST3.
CC MIM; 601408; --
CC GO; GO:0006323; P:DNA packaging; TAS.
CC InterPro; IPR005818; Histone H1/H5.
CC InterPro; IPR002717; MOZ_SAS.
CC InterPro; IPR001965; Znf_PHD.
CC Pfam; PF01853; MOZ_SAS; 1.
CC Pfam; PF00628; PHD; 2.
CC SMART; SM00526; H15; 1.
CC SMART; SM00249; PHD; 2.
CC PROSITE; PS01359; ZF_PHD_1; 1.
CC PROSITE; PS00016; ZF_PHD_2; 2.
CC Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
CC Nuclear protein.
KW ZN_FING 206 265 PHD-TYPE 1.
FT ZN_FING 259 313 PHD-TYPE 2.
FT DOMAIN 371 379 POLY-SER.
FT ZN_FING 538 560 C2HC-TYPE.
FT DOMAIN 788 801 POLY-GLU.
FT DOMAIN 989 995 POLY-GLU.
FT DOMAIN 1019 1026 POLY-ARG.
FT DOMAIN 1069 1078 POLY-GLU.
FT DOMAIN 1147 1150 POLY-LYS.
FT DOMAIN 1221 1242 GLU-RICH.
FT DOMAIN 1267 1302 GLU-RICH.
FT DOMAIN 1411 1414 POLY-GLU.
FT DOMAIN 1593 1597 POLY-SER.
FT DOMAIN 1643 1704 GLN/PRO-RICH.
FT DOMAIN 1897 1977 MET-RICH.
FT SITE 1546 1547 BREAKPOINT FOR TRANSLOCATION TO FORM
FT MOZ-CBP.
SQ SEQUENCE 2004 AA; 225054 MW; 9FFBAC3792854BA CRC64;

Query Match 2.4%; Score 322.5; DB 1; Length 2004;
Best Local Similarity 19.8%; Pred. No. 0.0013;
Matches 282; Conservative 199; Mismatches 582; Indels 359; Gaps 62;

QY 16 EPRYPHSLSPVQIARTHTDVGLLBYQHHSRDYASHLSPGSIQ-----PQRR 64
Db 801 EPQCRERELEISVGKSVSHENKQDSYSEKKEPVMAFVSTRLSKQVLPDLSLPANS 860
QY 65 RPSLLSEFQPGNRSQELHLRPESHSHYLPGLKSEMEFTESKPRLELLDPLLRPSPL 124
Db 861 QPSRRGRWGRKRRKTKRQFERFGDKSKLLLETSAPQEQYGCCEKSEATQEQYTES 920
QY 125 ATG--OPA--GSBDLTKDRLTKGLPVPSPPHDTPDELELVPPRLSKELIQNMDRVD 180
Db 921 VASEEQPSQDGKPDLPK--RLSEGVP-----WRGQLKKSPEAL--KRLTEGSESLP 970
QY 181 REITWVEQIISKKKQQLLEEAAPKEPEKPVSPPTTESKHSRLVQIITYDENRKAEEA 240
Db 971 RRYSEGDRAV--LRGFSSESEE-----EEPESPRSSP----- 1002
QY 241 AHRILEGLGQVELPLYNQPSDTRQVHENIKINQAMRKLLLYFKRNHARKQWKQK-FC 299
Db 1003 -----PILTKPT-----LKRKKPFLRRRRVRKRKHNSSVVT 1035

QY	300	QRDQLMEAL-----EKKVREIENPRRAKESKVREYVKOFFPIRKQELQERM	350
Db	:	: : : : : :	:
Db	1036	ETISETTEVLDEPFEDSDSERPWRL\$-PTEIDSEEEEDENLFP-----REYFRL	1088
QY	351	QSRVGORGSLSMSAARSHEVSEIIDGLSEQENLEKQMRQLAVIPMLPYDADOORIKFI	410
Db	:	: : : : : :	:
Db	1089	SSQDLRCOSSSKRSKBEE-----DEB\$D-----DADDTPI---1121	
QY	411	NNMGLMADPMKVVKORQVMNMHSEOKETFREKFMQHPK-----NFGIL	453
Db	:	: : : : : :	:
Db	1122	-----LKDVSLIRKRDVN\$PLETDTPLPKKKKGWPKGSRKPIHWKPGRKPQFKL	1175
QY	454	IASFUERKTVAECVLIVLYLTTKKNYKSILVR\$RRRGKSOQQOQQOQQOQQOQPMP	513
Db	:	: : : : : :	:
Db	1176	SREIMPVSQCA-CVI-----EPIV-SIPKAGR-KPKIQESEETVEPKEDMPLP	1220
QY	514	R\$SQBEKEKEKEAEABEKEPEVE-----NDKEDLLKEKTDOTSGDNDEKE	562
Db	:	: : : : : :	:
Db	1221	-----ERKEBEEMQAAEAEBEGEEDAASSEVPAA\$PADSNSPETETKEPEVEBEKP	1277
QY	563	AVASKGRKTANSQGRRKGRITRSMANEANSEBAITPQOSAELASWELNESSRWTEEMET	632
Db	:	: : : : : :	:
Db	1278	RYS\$EQRO\$EEOQ-----QELEEPEPE\$-BEDAAATAQNDHDDADEDDGHLES	1326
QY	623	AKIGLLEHGRNWSAJARMVGSKTYSQCKNFYNYKGRQNLEIILOQHKLXW-EKERNARR	681
Db	:	: : : : : :	:
Db	1327	TKKKELEOPTREDYKKEBPVQO-----ESF-----LDANNQ\$SREKIKOKEETELD	1372
QY	682	KKKKAP-----AAA\$E\$EAFPPVVDEEMEASGV\$NSEEMVEEABALHASGNEVPGECS	737
Db	:	: : : : : :	:
Db	1373	\$EEEOPSHDTSVVSQMAGS---EDDHED---SHTKEELIELKEE-----EIPHSEL	1421
QY	738	GPAUVNN\$DTESTPSPHTEAAKDTQGNGPKPPATLGADGPPGPPTPRRTGRAPTPT	797
Db	:	: : : : : :	:
Db	1422	LETVOAQVSLTOE\$SEHEGAYDCEE-----TLAACQTLOS\$YQDAEDPQMSWVEDC	1474
QY	798	PASEATGAPTPPPPAPPSP---SAPPPVVPKEK-----EETAAPAVEEGBEOK---	844
Db	:	: : : : : :	:
Db	1475	HASENSPIS\$VQ\$HPSQSVRSVSNPVPALESGYTOISPQ\$GSLAPSMMNETSPMD	1534
QY	845	-PPAAE--ELAVDTGKAEEPVKSECTE\$E\$E\$GPAKGDAAEAATAEGALKAEKKEGGSG	901
Db	:	: : : : : :	:
Db	1535	VPSVSDHSQQVVD\$GFDLG-\$IB\$STTENYEN\$SYDSTMGG\$ICGNSSQ\$SCSYGGLS	1593
QY	902	RATTAKSG---APQDS\$SATCSADEVDEAGGDKNRLL\$PR\$LLTPTGDP\$RANAS--	956
Db	:	: : : : : :	:
Db	1594	\$SSSLTOS\$CVVTQOMASMG\$SCSMQO\$SSVPAANC\$IK\$PQ\$CVVE---RP\$NQ\$QOQ	1650
QY	957	-----POKPLDLKQLKARAAAIPIIOVTKVHEPREDAATKAPAPBPQNLQPEDSA	1011
Db	:	: : : : : :	:
Db	1651	PPPPPPQOPPPPOPAPAQP PPPPQ-----QOPQOQOPQOQPPPPPPPPQOQPLSOC	1705
QY	1012	PQO\$G\$PRGKSRP-----APPADKEAFAEAOKLPQDPPCWTSGLP---FVPVPREVI-	1063
Db	:	: : : : : :	:
Db	1706	\$MNN\$FTP-----APMIMEIPESGSTGNISIYERIPGDFGAGSYSQPSATFSIAKLQILT	1760
QY	1064	-----KASHAPDPP\$AF\$YA-----PFCHEPLGLHDTARPVLPRPPTISN	1104
Db	:	: : : : : :	:
Db	1761	NTIMDPHAMYP\$H\$PAVTSVATS\$VSI\$NTGLAQLAP\$HFL-----AGTPOAQATMT	1811
QY	1105	PPPLISSAK---HPSVLIEROIGAISOQMSVOLHVHYSEHAKAPGVPTWGLPLPMDPKKL	1161
Db	:	: : : : : :	:
Db	1812	PPPNLASTTNLTSPLLQCWN\$ATNIG-----IPTHORLOQOM-PVKGHISIR\$K\$APL	1864
QY	1162	AP\$GVKQEOQL\$PRG-----QAGPPESLGVP\$TAQ\$ASVLRGTALGSVPG\$ITGI\$PST	1215
Db	:	: : : : : :	:
Db	1865	-P\$AAAHQOOLYGR\$PSAVAMQAGP-----RALAVORGMNMGMVNLMP\$PAYNVNSM	1914
QY	1216	RVP\$SD\$AI-TYR-----GSITHGT\$ADVLYKGTITRIIGED\$PSRLDRG\$EDSL----	1263
Db	:	: : : : : :	:
Db	1915	NMN\$TLNAMNSYRMTQPMN\$SYHSNFA---YNNQTAQY-----PQMOMGMG\$QAYTQO	1966

"Restricted homology between human alpha 1 type IV and other procollagen chains.";
Proc. Natl. Acad. Sci. U.S.A. 82:3649-3653(1985).
[8]
SEQUENCE OF 1-28 FROM N.A.
MEDLINE=89034231; PubMed=3182844;
RA Soaninen K., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
[9]
SEQUENCE OF 1441-1669, AND DISULFIDE BONDS.
RC TISSUE=Placenta;
RX MEDLINE=89005112; PubMed=2844531;
RA Steibold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
CC -!- FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC -!- SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV) -
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC -!- DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC -!- PTM: Lysines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in all cases and bind carbohydrates.
CC -!- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -!- PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M26576; AAA53098.1; JOINED.
DR EMBL; J04217; AAA53098.1; JOINED.
DR EMBL; M26550; AAA53098.1; JOINED.
DR EMBL; M26540; AAA53098.1; JOINED.
DR EMBL; M26542; AAA53098.1; JOINED.
DR EMBL; M26543; AAA53098.1; JOINED.
DR EMBL; M26544; AAA53098.1; JOINED.
DR EMBL; M26545; AAA53098.1; JOINED.
DR EMBL; M26546; AAA53098.1; JOINED.
DR EMBL; M26547; AAA53098.1; JOINED.
DR EMBL; M26537; AAA53098.1; JOINED.
DR EMBL; M26538; AAA53098.1; JOINED.
DR EMBL; M26539; AAA53098.1; JOINED.
DR EMBL; M26548; AAA53098.1; JOINED.
DR EMBL; M26549; AAA53098.1; JOINED.
DR EMBL; M26551; AAA53098.1; JOINED.
DR EMBL; M26552; AAA53098.1; JOINED.
DR EMBL; M26553; AAA53098.1; JOINED.
DR EMBL; M26554; AAA53098.1; JOINED.
DR EMBL; M26555; AAA53098.1; JOINED.
DR EMBL; M26556; AAA53098.1; JOINED.
DR EMBL; M26557; AAA53098.1; JOINED.
DR EMBL; M26539; AAA53098.1; JOINED.
DR EMBL; M26558; AAA53098.1; JOINED.
DR EMBL; M26559; AAA53098.1; JOINED.

DR EMBL; M26560; AAA53098.1; JOINED.
DR EMBL; M26561; AAA53098.1; JOINED.
DR EMBL; M26562; AAA53098.1; JOINED.
DR EMBL; M26536; AAA53098.1; JOINED.
DR EMBL; M26563; AAA53098.1; JOINED.
DR EMBL; M26541; AAA53098.1; JOINED.
DR EMBL; M26564; AAA53098.1; JOINED.
DR EMBL; M26565; AAA53098.1; JOINED.
DR EMBL; M26566; AAA53098.1; JOINED.
DR EMBL; M26567; AAA53098.1; JOINED.
DR EMBL; M26568; AAA53098.1; JOINED.
DR EMBL; M26569; AAA53098.1; JOINED.
DR EMBL; M26570; AAA53098.1; JOINED.
DR EMBL; M26571; AAA53098.1; JOINED.
DR EMBL; M26572; AAA53098.1; JOINED.
DR EMBL; M26573; AAA53098.1; JOINED.
DR EMBL; M26574; AAA53098.1; JOINED.
DR EMBL; M26575; AAA53098.1; JOINED.
DR EMBL; Y00706; CAA68698.1; -.
DR EMBL; X05561; CAA23075.1; -.
DR EMBL; M10940; AAA52006.1; -.
DR EMBL; M11315; AAA52042.1; -.
DR PUR; S16876; CGHU4B.
DR Genew; HGNC:2202; COL4A1.
DR MIM; 120130; -.
DR InterPro; IPR008161; C1g helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR001442; Procollagn4_C.
DR Pfam; PF01413; C4; 2.
DR Pfam; PF01391; Collagen; 24.
DR ProDom; PD000007; C1g helix; 6.
DR ProDom; PD003923; Procollagn4; 1.
DR SMART; SM00111; C4; 2.
DR Repeat; Hydroxylation; Glycoprotein; Signal.
FT SIGNAL 1 27
FT PROPEP 28 272
FT CHAIN 173 1669
FT DOMAIN 173 1440
FT DOMAIN 1441 1669
FT CARBOHYD 126 126
FT DISULFID 146 1551
FT DISULFID 1493 1548
FT DISULFID 1505 1511
FT DISULFID 1570 1665
FT DISULFID 1604 1662
FT DISULFID 1616 1622
FT CONFLICT 237 238
FT CONFLICT 241 241
FT CONFLICT 319 319
FT CONFLICT 719 719
FT CONFLICT 837 837
FT CONFLICT 842 842
FT CONFLICT 896 896
FT CONFLICT 904 904
FT CONFLICT 914 914
FT CONFLICT 998 998
FT CONFLICT 1010 1010
FT CONFLICT 1012 1012
FT CONFLICT 1358 1358
SQ SEQUENCE 1669 AA; 160611 MW; 3BBBA6DFFB9BA84 CRC64;
SG -> K (IN REF. 4).
G -> K (IN REF. 4).
Q -> A (IN REF. 3).
N -> D (IN REF. 5).
D -> Y (IN REF. 5).
K -> P (IN REF. 5).
V -> W (IN REF. 2).
E -> Q (IN REF. 5).
S -> K (IN REF. 5).
K -> P (IN REF. 5).
S -> K (IN REF. 5).
E -> Q (IN REF. 5).

Query Match 2.4%; Score 320.5; DB 1; Length 1669;
Best Local Similarity 21.0%; Pred. No. 0.0013;

Matches 415; Conservative 133; Mismatches 750; Indels 677; Gaps 100;
Qy 728 GNEVPRGECGSPATVNNSSDTESIPSHTRAAK-DTQNG-----PKPPATLGADGPPPGP 782
Db 64 GMOGEGP-QGPP--GQKGTGEGPLGTRGPGCASGYPGNGLPGIPGQDG-PPGP 119
Qy 783 PTPP-----RTSRAPTEPTASEATGATP----- 808
Db 120 FGIPGCGNGTGERGRLGPGPLGPGFAGNPGPPGLPGMGKDPGEILGHVPMLLKGERGPPG 179

QY 809 -----PPAPSPSAPPV-----PKKEKEETAAP 835
Db 180 IPTGPPGGLPGLGVPPGPGTGGPPGPGGKGMGLSFGQPKGDKGDQVSGP 239
QY 836 PVEGEQKPPAAEELAVDTKAEPEVKSECTEAEAG-----PAKGDAAEAATAAGAL 891
Db 240 PGVGGQAQ-----KGFATKGEKQKGFPGQMPGVCEKGEKPKGPRG-- 287
QY 892 KAEKKEGSGRATTAKSGAPQSDSSATCSADEVDEAGGDKNRLSP-RPSLLTPTGD 950
Db 288 -----KPKGDGKGEKSGPGPGPGYPLIGR-----QGPQGEKGEAGPGPGPIVIGT- 338
QY 951 PRANASQKPLDLKQLKORAAAPPIQVTKVHEPPREDAAPTKAPAPPAPPPQNQLQPSD 1010
Db 339 -----PLGEK-----GERYPTGPRGEPGPGKPGFLPG 368
QY 1011 APOQSSPGKRSRAPPADK-----EAPAAEAKLPGDPPFCWTSGLPFPVPPREVIKA 1065
Db 369 QPGPPGLPVGAQAGFPGEERGEKGDGFGPGTS-LPG--PSGRDGLPG----- 415
QY 1066 SPHAPDPSAFSYAPPCHPLPLGLHDTARVLPRLPTISNPPPL-----ISSAKH 1114
Db 416 -PGSP-----GVTNGIVECQPGPDGPGPGIQQPGFPGISGEKQKG 464
QY 1115 PSVLRIQIGAI-----SQGMSVQLHVYSEHAKA-----PVGPTVMTGLPLPMDPKKLAPFS 1165
Db 465 ESCLICIDIGYRPGPGPGPGPGIQQPGQKAGDRGLPGRDGVAGVPGPGTGLIQP 524
QY 1166 GVKQE-----QLSPQAGPESIGVTPAQEASVLRGTALGSPVGGSIITKIGPSTRVPS 1219
Db 525 GAKGEPEGYFDLRUKGDKDGFPGQPG-----MPGRA-GS-PGRDGHPLCPKPK-GS 575
QY 1220 DSAITYRGSITHTPADVLYKTIITRIIGEDSPSLDRGREDSLPKGHVIE-----GKK 1274
Db 576 PGSVGLKGE--RGPPGVGFGS-----RGDTGPPGPGPGGAPGIDCK 617
QY 1275 GHVLSYEGMSVTCSEKDRSSGPPHETAAPKRTYDMMGRVGRATISSASIEGLMG-- 1332
Db 618 G-----QAGFPFGPGSLPGPKGEP-----GKIVPLPGPGAGSLPGSP 657
QY 1333 -----RAIPERHSPHLKEQHIGRSITO-----GIPRSYVBAQEDYLREAKLKRE 1381
Db 658 GPPGQGDGPGTGRGLPGE-----KGAVQPGIGPGPGPGKGVDPGPD-----MGPP 710
QY 1382 GTTPPPPP-----PSRDLTEAYTQ-----ALGPLKLPHEGLVATVKEAGR-SIHEIPREE- 1432
Db 711 GTPGRPGFNGLPGNPGVQKGEKPGVGLPLGLPLGLPGIPGTPGKSGSIGVPGVPGEHG 770
QY 1433 -----LRHTRELPLAPRLKEGSI--TOGTPLKVDYTGASTTGSK-----KHDVRLIG 1478
Db 771 AIGPPGLQIGEPGPGPL-----GSVSGPGVP-----GIGPPGARGPPGQPGPLSG 820
QY 1479 SPG-----RTFPVPHPLDVMA-----DARALERACVEESLKSRTPGTASSSGSITARGAPVI 1529
Db 821 PPGIKGEKGFPGFLDMPGPKGDKAGQLPGITGQSLPCLPGLPQOQAPGIPGPGSKGE 880
QY 1530 VPGLKPRQS-----PLYEDHGAFFAGHLPRGSPVTWRE-----PTP 1567
Db 881 MGVMGTQPGQSGPGVAGAPLPGKGDHGFPGSSG-PRGDPGLKDKGVDVGLPKPKGSM 939
QY 1568 RLOEGSLSSSKASQDRK-----LTSTPREIAK-----SPHSTVPEHHH--- 1606
Db 940 KVDGSMKMGKGDQGEKQIGIPGKSGRDPGTGPGVPGKQAGQPGQPGKDPGIGSG 999
QY 1607 -----PISPYEHLRLGVSGVDLYRSHIPLAFDPTSIPIRGIPLDAAA 1647
Db 1000 TPAGAPLPGPKSGVGMGLPFTGPE---KGVPGIP-----GPGQSP-GLPGDKGA 1045
QY 1648 AYXLPRLHAPNTYPLHYPPVLYRNGYDDTAALENRQIINDYITSQMHNTATANAQRA 1707
Db 1046 KGEKQAGPPGIGIPGL-----RG-----BKG 1067

QY 1708 DM-LRGL--SPRESSLALNYAAGPRGIIDLSOVPHLPVL-----VPTEGTPTATMD 1756
Db 1068 DOGIAGPFGSPGEK-----GEKSGIGIPGMPGSGPGLKSGPGSVGYPGSGPLPGBKGD 1119
QY 1757 RLAVLPTAPQ-PFSSRISSPLSPGPGTHLTKPTTTSSSERERDRDRDREREKSIL 1815
Db 1120 K--GLPGLDGIPIGVKGEAGLFGTGF-----PTGPAGQKGEPGSGDIPGSAKEGEPGL 1170
QY 1816 TSTTTVEHAPIWRPGTBTQSSSGSSSG-----GGGSSSRPASHSHAHQHSPISTPTODA 1870
Db 1171 PGRGF-----PGFPAKGDGKSGKEVFPGLAGSPGIPGSGKEGFGWPGPGQPG 1222
QY 1871 LQRPSPVLHNT-----GMKGIIATAPEPSKPTVLRSTSTSSVPRPAATPPATHCPGLG 1923
Db 1223 LPSFG--HATEGPKGDRGPGQ-----QPGLPGL-----PGMPGPG--LPG 1260
QY 1924 TLDGVYPTLMEPVLLPKEAPRVARPERPRADTGHAFKAPPARSGLEPASSPSKSGSEPRP 1983
Db 1261 -IDGVKDKGNP-----GWFGAPGVPGKGDGPG--FGMPGI--GGSPGIGTSKGDMPGP 1310
QY 1984 LVPPVSGHATARTP-----AKNLAPHIASPDPPAPPASASDPHREKTQSKP 2030
Db 1311 GVPGFQPGKGLPGLQIKGQDQGVPGAKGL-----PCPPGPG--P 1351
QY 2031 FSTOELELRSLGYHSSYSPEGVEPVSSPSLTHDKGLPKHLELDKSHLEGEURPK- 2089
Db 1352 YDIIKGE--PGLPG-----PEG-----PPGLKGLGLP-----GPKG 1381
QY 2090 QPGFVKLGEAAHLHLPLPESQSSSPLLQTAGVKGHQVVTLLAQHI SEVITODYTR 2149
Db 1382 QQGVTLGVG-----IP--GPGIIPGFGAPQKQK----- 1408
QY 2150 HHQQLSAPLPAPLYSPGASCVDLRLRPPSDLYLPDPDHGAPARGSPHSGKRSPEP 2209
Db 1409 -----EMGPAGTGTGPRGPGP-----PGPD-----GLPGSMGPPGPTPSV 1442
QY 2210 NKTSVLGGEDGIEPVSPPEGMTEPCHSRSAVYLLYRDG-EQTEPSRMGSKSP--GNYS 2266
Db 1443 DHGFLVTRHSQIIDDQPCSGKILYHG-----YSLLYVOGNERAHQDGLTAGSCLRKFS 1498
QY 2267 QPPAFPSKLTESNAMSVMKSKQEIINKLTHNRNPEYNISQPGTEIFNMPAITGTGLMT 2326
Db 1499 TWPFLFC-----NINNVCFASRNDYSYWLSTPEPMPSMAPIITGENIRP 1543
QY 2327 YRSQAQVEHASTNMGLEAIRKALMGKYDOWEESPLSANAPNPLNASASLPAAMPITAA 2386
Db 1544 FISRCVACBAP-----MVNAVHSQTIQIIP-CPGSMSSLWIGYSFVM----- 1585
QY 2387 DGRSDHTLTPGGGGKAKVSGRPSRRKAKSPAPGLASGD-RPPSVSVSHSEDCN 2440
Db 1586 -----HTSAGAEKSGOALAS-----PGSCLEEFRSAPFIECHGRGTCN 1623

RESULT 89

CA21-CHICK

ID CA21-CHICK STANDARD; PRT; 1362 AA.
AC P02467; P87491; Q90758; Q90792; Q90795; Q90797; Q92014;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Collagen alpha 2(I) chain precursor (fragments).
GN COLIA2.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE OF 1-245; 262-448 AND 466-1362 FROM N.A.
RX MEDLINE=86185168; PubMed=368961;
RA Boedtker H., Finer M., Aho S.;
RT "The structure of the chicken alpha 2 collagen gene.";
RL Ann. N.Y. Acad. Sci. 460:95-116(1985).

RT SEQUENCE OF 1-89 FROM N.A.
RP MEDLINE=83246518; PubMed=6135195;
RA Tate V.E., Finer M.H., Boedtker H., Doty P.;
RT "Chick pro alpha 2 (I) collagen gene: exon location and coding
RT potential for the prepropeptide."
RL Nucleic Acids Res. 11:91-104 (1983).
[3]
RT SEQUENCE OF 1-14 FROM N.A.
RP MEDLINE=82060240; PubMed=6946474;
RA Vogeli G., Ohkubo H., Sobel M.E., Yamada Y., Pastan I.,
RA de Crombrughe B.;
RT "Structure of the promoter for chicken alpha 2 type I collagen gene."
RL Proc. Natl. Acad. Sci. U.S.A. 78:5334-5338 (1981).
[4]
RT SEQUENCE OF 1-33 FROM N.A.
RP MEDLINE=84297217; PubMed=64731103;
RA Aho S., Tate V.E., Boedtker H.;
RT "Location of the 11 bp exon in the chicken pro alpha 2 (I) collagen
RT gene."
RL Nucleic Acids Res. 12:6117-6125 (1984).
[5]
RT SEQUENCE OF 1-79 FROM N.A.
RP MEDLINE=88056316; PubMed=3678934;
RA Finer M.H., Boedtker H., Doty P.;
RT "Construction and characterization of cDNA clones encoding the 5' end
RT of the chicken pro alpha 1 (I) collagen mRNA."
RL Gene 56:71-78 (1987).
[6]
RT SEQUENCE OF 78-92.
RP TISSUE=Skin;
RX MEDLINE=71115216; PubMed=5544653;
RA Highberger J.H., Kang A.H., Gross J.;
RT "Comparative studies on the amino acid sequence of the alpha 2-CB2
RT peptides from chick and rat skin collagens."
RL Biochemistry 10:610-616 (1971).
[7]
RT SEQUENCE OF 74-91; 263-448 AND 1088-1169 FROM N.A.
RP MEDLINE=82059081; PubMed=6272119;
RA Wozney J., Hanahan D., Tate V.E., Boedtker H., Doty P.;
RT "Structure of the pro alpha 2 (I) collagen gene."
RL Nature 294:129-135 (1981).
[8]
RT SEQUENCE OF 78-92.
RP TISSUE=Skin;
RX MEDLINE=70131186; PubMed=4313735;
RA Kang A.H., Gross J.;
RT "Amino acid sequence of cyanogen bromide peptides from the amino-
RT terminal region of chick skin collagen."
RL Biochemistry 9:796-804 (1970).
[9]
RT SEQUENCE OF 78-92 AND 415-448.
RP TISSUE=Skin;
RX MEDLINE=69285369; PubMed=5809220;
RA Kang A.H., Igarashi S., Gross J.;
RT "Characterization of the cyanogen bromide peptides from the alpha-2
RT chain of chick skin collagen."
RL Biochemistry 8:3200-3204 (1969).
[10]
RT SEQUENCE OF 78-92 AND 415-448.
RP TISSUE=Bone;
RX MEDLINE=69206882; PubMed=5785233;
RA Lane J.M., Miller E.J.;
RT "Isolation and characterization of the peptides derived from the
RT alpha 2 chain of chick bone collagen after cyanogen bromide
RT cleavage."
RL Biochemistry 8:2134-2139 (1969).
[11]
RT SEQUENCE OF 566-587 FROM N.A.
RP MEDLINE=79074829; PubMed=364479;
RA Lehrach H., Frischauf A.-M., Hanahan D., Wozney J., Fuller F.,
RA Crkvenjakov R., Boedtker H., Doty P.;
RT "Construction and characterization of a 2.5-kilobase procollagen

clone.";
RT Proc. Natl. Acad. Sci. U.S.A. 75:5417-5421 (1978).
[12]
RT SEQUENCE OF 902-1362 FROM N.A.
RP MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1 (I) and pro-alpha 2 (I) collagen messenger ribonucleic acids
RT including the carboxy-terminal propeptide sequences."
RL Biochemistry 20:996-1006 (1981).
[13]
RT SEQUENCE OF 998-1169 AND 1234-1362 FROM N.A.
RP MEDLINE=81264246; PubMed=6267043;
RA Dickson L.A., Ninomiya Y., Bernard M.P., Pesciotta D.M., Parsons J.,
RA Green G., Eikenberry E.F., de Crombrughe B., Vogeli G., Pastan I.,
RA Fietzek P.P., Olsen B.R.;
RT "The exon/intron structure of the 3'-region of the pro alpha 2 (I)
RT collagen gene."
RL J. Biol. Chem. 256:8407-8415 (1981).
[14]
RT SEQUENCE OF 932-954 AND 968-980 FROM N.A.
RP MEDLINE=81064671; PubMed=6159982;
RA Avvedimento V.E., Vogeli G., Yamada Y., Maizel J.V. Jr., Pastan I.,
RA de Crombrughe B.;
RT "Correlation between splicing sites within an intron and their
RT sequence complementarity with U1 RNA."
RL Cell 21:689-696 (1980).
[15]
RT SEQUENCE OF 126-161; 467-517 AND 926-954 FROM N.A.
RP MEDLINE=8112157; PubMed=7460017;
RA Yamada Y., Avvedimento V.E., Mudryj M., Ohkubo H., Vogeli G.,
RA Irani M., Pastan I., de Crombrughe B.;
RT "The collagen gene: evidence for its evolutionary assembly by
RT amplification of a DNA segment containing an exon of 54 bp."
RL Cell 22:887-892 (1980).
CC -1- FUNCTION: Type I collagen is a member of group I collagen
CC (fibrillar forming collagen).
CC -1- SUBUNIT: Trimer of one alpha 2(I) and two alpha 1(I) chains.
CC -1- TISSUE SPECIFICITY: Forms the fibrils of tendon, ligaments and
CC bones. In bones the fibrils are mineralized with calcium
CC hydroxyapatite.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; M25963; AAA69960.1; JOINED.
DR EMBL; M25956; AAA69960.1; JOINED.
DR EMBL; M25959; AAA69960.1; JOINED.
DR EMBL; M25961; AAA69960.1; JOINED.
DR EMBL; M25962; AAA69960.1; JOINED.
DR EMBL; M25965; AAA69961.1; JOINED.
DR EMBL; M25964; AAA69961.1; JOINED.
DR EMBL; M25984; AAA69962.1; JOINED.
DR EMBL; M25957; AAA69962.1; JOINED.
DR EMBL; M25966; AAA69962.1; JOINED.
DR EMBL; M25967; AAA69962.1; JOINED.
DR EMBL; M25969; AAA69962.1; JOINED.
DR EMBL; M25970; AAA69962.1; JOINED.
DR EMBL; M25971; AAA69962.1; JOINED.
DR EMBL; M25972; AAA69962.1; JOINED.
DR EMBL; M25973; AAA69962.1; JOINED.
DR EMBL; M25974; AAA69962.1; JOINED.
DR EMBL; M25976; AAA69962.1; JOINED.
DR EMBL; M25977; AAA69962.1; JOINED.
DR EMBL; M25978; AAA69962.1; JOINED.
DR EMBL; M25979; AAA69962.1; JOINED.

Db 107 QGVQQLGLELRPRVFLYEDQGRPOASQPIFRGLSLADGKWHVAVAKGQSVT- 162
Qy 652 FYFNKKR-----ONLDEILQOHKLKWEKERNARRKKKAPAAEAEAFPPVVEDEM 705
Db 163 LIVDCKRVRPLRSPVHVDTHGVVIFGAH-----ILDEYF 201
Qy 706 EASGVSGNEBEM--VEBAEALHASGN-----EVRPGECGPATVNNSSDITESIPSPHTEAA 759
Db 202 E-----GDVQELLVWPGVQAAAYSCQKOLECEREQDGPOT-----QKPH- 242
Qy 760 KDTQNGCP-KPPATILGADGPPGPPPPRRTSRAPTEPTASATGAPTTPPAPPSPSAP 818
Db 243 --RAQSPKKEBAPLARHKP-QSOEPQOKPTESLYDYEPYVDVMTGTAPDYQYPTPG-- 297
Qy 819 PPVVPKEKEEETAAPPVBEGBE--QKPPAAEELAVD--TCKAEPEV-----XS 864
Db 298 -----EEEGVLESPPFLFEEBQTDLOVSTADSFQAEYEGEGTDSFAGFYDYTYGYGD 352
Qy 865 ECTEEAEGBAKGDAAEAPATAEG--ALKAEK-----KEGSGRATTAKSGA 911
Db 353 DYREETELGALSAETAHSAVAHAPRGLKGEKEBPAVLEPGMFVEGPPGPEGPAGLAGP 412
Qy 912 PQSDSSATCSADEVDEAGCDKXRLLSRPSLLTPTGDP-----RANASPOKPL 961
Db 413 PGIQGNPQGVG-----DPGERGPPGR--AGLPGSDGPPGPGTSLMLPFRFGSGGDKGPV 466
Qy 962 DLKQLKQRAAAIPIPIQVTKVHPPPREDAAATKPAAPPAPPPQNLOPESD--APQOPGSSP 1019
Db 467 VAAQEAQAAILQQAARLA-LRGPPGPMGYTGPRGLGQPGSLKGESGDLGPQ-----GP 521
Qy 1020 RGKSRSPAPADK-----EAPAAEAQKLPDPPC-----WTSGLP-----F 1055
Db 522 RGQQLGTGPPGKAGRRAGADGARGMPGEMKGDGRFDGLPLPGEKGQRGDTGAQGL 581
Qy 1056 PVPREVIKASHPADPSAFSVAPCHPLPLGLHDTARPVLPR-PPTISNPPPLISSAKH 1114
Db 582 PGPPGE--DGERGDDGEIGRGLFOESGFRGL-----LGPKGPPGPGPPGV----- 626
Qy 1115 PSVLEROIGAIQSGMSVQLHVPVSEHAKAPGVFTMGLPLPMDPKLAPFSGVQEQQLSP 1174
Db 627 -----RGMD-----GPHGP-----KGS LGP 641
Qy 1175 RQOAGPPESLGVPTAQEASVLRGTALG-----SVPGSITKGIPS 1214
Db 642 QSEPPPGQGTGPAQGLPGPQG-AIGPHGEKARGKPLGPMFPGSDGLPGHPKEGPPG 700
Qy 1215 TR---VPSDSAITYRGSITHGTTPADVLYKGTITRIIGEDSPSLDRGREDSLP--KGHI 1269
Db 701 TKGNQGPSPG---QGPLGYGPRGVKGVGIRGLKGHKG---EKG-EDGFPFGFGDIG 751
Qy 1270 YEGKKGHLVSEGMVSTQCSKEDG-----RSSGPPHETAAPKRTYDME----- 1315
Db 752 VKGDRGEV-----GVPSRGEDGPEGKGRGTGP---TGDPGPTGLMGEKGLGVGLP 801
Qy 1316 ---GRVG-----RAISSAIEGLMGRAIPPERHSPHLLKEQHHIRGSI----- 1355
Db 802 GYFGQGGKSLGFPFGFAGSEKARGLSGKSGRGERGTGPRQGRGPRGATGKSGAK 861
Qy 1356 -----TGIPRSTVEAQEDYLRKALKKREGTP-----PPPPPSRDLTEAYK 1398
Db 862 GTSGDGPHGPPGERGLPGPQ-----GPNFPGPKGPPGAGKDLGPH 906
Qy 1399 TQ-----ALGPLKLPKPAHEGLVATVKEAGRSIHIPREELRHT--PELPLAPRPLKEG 1449
Db 907 QORBEVGGQTKGP-----PGPPGVVPGGTAGES---GPMGERGHSQFPGPPGQGLPGT 959
Qy 1450 SITQGTPLKYDTGASTGSKKHVRSLLIGSPG-----RTEPPVHLDVWADARALERA 1502
Db 960 SKEGT--KGDGPP-----GAFKDGKAGLAGRFGFQERGLPCTAGCGPLKG- 1003
Qy 1503 CYESLSKRPCTASSG--GSIARGAPVIVPELKGKPRQSPLTVEDHGAPPAGHLPRGSPV 1560

Db 1004 --NECPAGPPGAPGAPGGERGAAGSGGPTGPP--GRPG-----PQGP 1042
Qy 1561 TWREPTPLQOBSLSSSSKASQDRKLTSTPRBIAKSPHSTVPEHHPHPTISPYEHLRLRGVSG 1620
Db 1043 AAGEKGVPGKXGPIG-----PTGRDGVQGPVCL-----PGAPG-----PGVAG 1081
Qy 1621 VDLRYSHPLAFADPTISPRGIPLDAAAAYLPRHLANPTYPHLYPPYLIRGYDPTAALE 1680
Db 1082 EDGDKGEV-----GDPGQKGTGKNKGEGHPGPPGP-IGPV---GQGAAG 1125
Qy 1681 NRQTIINDYITSQQMHNTATAMARADMLRGLSPRESSLALNYAAGPRGIIDLSQVPHL 1740
Db 1126 GEPG-----ARGPQGHFG-----AKGDETRGFNGPPGPIGLQQLPGPSG--EKGETGDG 1173
Qy 1741 PVLVPPPTGTATAMDRLAYLPTAPQPFSSRHSSPLSPG-----PTHLTPTTTSSS 1794
Db 1174 GPMGPPGPPGPRG-----PAGPNGADGPGQ---SPGCVGNLPGPEKGEFEGSGSP 1221
Qy 1795 ERERDRDRDRDREREKSIILSTTTTVEHAPIWRPGTEQSSGSSGSGGGG--GSSSRPA 1852
Db 1222 GVQGEVGVKGRGERGEG--ESQAGEAGP---PGPKGPTGDNKPKGNPGVPVFGDPG 1276
Qy 1853 SHSHAHQSPISPRTDALQORPSVLHNTGMKGLITAVEPSKPTVLRLSTSTSPVPAAT 1912
Db 1277 PGGA-----GPRQDGAK-----GDRGEDG-----EPGQSGPPTGENGPPGLGK 1319
Qy 1913 FPPA-THCPIGLTLDGVYPTLMBPVLLPKAPRVARPERPRADTGHAFIAKPPARSGL- 1970
Db 1320 RGPAGTGPPEG--ROGEKGAKGDPGAV--GAPKGTGP-----VGPAGLAGKPGDGLRG 1369
Qy 1971 -PASSPSKG-----SEPRPLVPP-----VSGH-----ATIARTPAKNLAP 2004
Db 1370 LPSGVQOQGRPGATGQAGPPGVPPQLPGLRDAGAKGKHPGLIGLIGTGEQGEKG 1429
Qy 2005 HHASPPDPAPASADPHREKTSQSFISQELRLSLYHSGSSYSPGVEPVSPVSSPSL 2064
Db 1430 DRGLPGPQGSF-----ELDKSHLSEGLRKPQGPVKLGEEAAHLPHLRPLPESOP--- 2114
Qy 2065 THDKGLPKHLE-----ELDKSHLSEGLRKPQGPVKLGEEAAHLPHLRPLPESOP--- 2114
Db 1464 PGPSG-PKANGATGAPGKGEKVGP--PGHPGP--PGEV-----IQPLIQMPKKT 1512
Qy 2115 ---SSPLLQT-----APGVKHQHVWTLAQHISBVI-TQDYTRHHPOOLSAPLA 2161
Db 1513 RRSVDGSKLIQDEEAVPTGAGPSA-----GLEIFGSLDSLRREEIQMRR--PA 1561
Qy 2162 PLYSFPAGSCVLDLRRP--PSDLYLPPDPHG----- 2191
Db 1562 GTQDSPARTQDILKLCHPELDPGEYVWDPNQCARDAFVFCNFTAGGETCVTPRDDVTQ 1621
Qy 2192 ---APARGSP-----HSR-----GKRSPEPNKTSVLGGGEGDIEPVSP 2227
Db 1622 FSVYDSEGSFVGVVQLTFLRLLSVAHQDVSPVCSGVSDGPK--LRGANEDELSPETS 1679
Qy 2228 PGMTEPHGSRSAVPLLYRDGEQTEPSR 2256
Db 1680 P-----YVKEFRDCQCTQGR 1695

RESULT 91

GPI_CHLRE

ID GPI_CHLRE STANDARD; PRT; 555 AA.

AC Q9FP06; Q03927;

DT 28-FEB-2003 (Rel. 41, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Vegetative cell wall protein gpi precursor (Hydroxyproline-rich

DE glycoprotein 1).

GN GPI.

OS Chlamydomonas reinhardtii.

OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;

OC Chlamydomonadaceae; Chlamydomonas.

OX NCBI_TaxID=3055;

RNA SEQUENCE FROM N.A.
MEDLINE=21159092; PubMed=11258910;
Ferris P.J., Woessner J.P., Waffenschmidt S., Kilz S., Drees J.,
Goodenough U.W.;
"Glycosylated polyproline II rods-with-kinks as a structural motif in
plant hydroxyproline-rich glycoproteins";
Biochemistry 40:2978-2987(2001).
[2]
PARTIAL PRELIMINARY SEQUENCE FROM N.A.
MEDLINE=91017504; PubMed=1699225;
Adair W.S., Apt K.E.;
"Cell wall regeneration in Chlamydomonas: accumulation of mRNAs
encoding cell wall hydroxyproline-rich glycoproteins";
Proc. Natl. Acad. Sci. U.S.A. 87:7355-7359(1990).
-1- FUNCTION: Major component of the outer cell wall W6 (crystalline)
layer.
-1- SUBUNIT: Associates with GP2 and GP3.
-1- PTM: N-glycosylated and O-glycosylated.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; AF309494; AAC45420.1; -
EMBL; M58496; AAA69706.1; ALT_SEQ.
Glycosylated; Q9P06; -
InterPro; IPR003882; Pistil_extensin.
PRINTS; PR01218; PSTEXTENSIN.
KW Glycoprotein; Repeat; Signal.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 555 VEGETATIVE CELL WALL PROTEIN GP1.
FT DOMAIN 40 339 49 X 5 AA APPROXIMATE PPSXP REPEATS.
FT DOMAIN 259 279 POLY-PRO.
FT CARBOHYD 399 399 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 455 455 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 493 493 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 555 AA; 54219 MW; 6A584A90465502F5 CRC64;
Query Match 2.4%; Score 317; DB 1; Length 555;
Best Local Similarity 24.4%; Pred. No. 0.00059;
Matches 149; Conservative 39; Mismatches 277; Indels 146; Gaps 21;
QY 717 MVEAEALHASGNVPRGECGPGATVNVNSDTEIPSPHTEAKDTQNGKPKPATIGAD 776
DB 18 MVVLAFAVANAQCVPGGIFNCPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 77
QY 777 GPPGPPPTP---PRRTSRAPIEPTPASEATGATPP-PAPPPSPAP-----PPVVPKE 825
DB 78 PPSGAPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSPAPPSP 137
QY 826 EKEEETAAAPVVERGEQKPPAAELAVDTGKAEPEVKSECTEAEAGKAGKDAEAAEA 885
DB 138 SPSPAPPPLPPSPAPPSPSPVPVP-----SPSPVPSPAPPSPPTPPSPVPVPSPA 190
QY 886 TAEGALKAEKGGSGRATTAKSSGAPQDSSTATCSADEVDEAGGDKRLLSPRSL 945
DB 191 PPSAPPVPVPSAPPSPAPPVPVPSAPPSPAPPSPAPPSP-----PSPA 232
QY 946 TPTGDRPRANASPKQLDLKQLKQRAAIPPIQVTKVHPREDAAPTKAPPPPPQNL 1005
DB 233 PPSGSPAPPSPVPVP-----SPAPSPAPPSPKPP-----APPPPPPPPPPPPPPP 279
QY 1006 QPESDAPQQRSGSRGKSRPAPPADKEAFAAEAQKLGPPGPPCWTSGLPFPVPPREV- 1064
DB 280 FP-ANTWMP--SPSPSPSPAPT-----PPTPP--SPSPSPVPVPSAPVP 322
QY 1065 ASHAPDPSAFSVAPPGHPLGLHDTARP-VLPRPTTINPPPLISSAKHPSVLERQIG 1123

Db 323 PSPAPPSPAP---SPPPSPAP-----PTPSPSPSPSPSPSPSPSPSPSPSPSP 371
QY 1124 AISQMSVQLHVPYSEHAAPVGPVTMGLPLPMDPKKLAPFSGVKQEQLSRPGAGPPES 1183
Db 372 -----SPSPKPSPPSPVAVKLVWADDAIAFDLLNGTSTRPGSASRVWGEPI 417
QY 1184 LGV-----PTAQEASVLRGTALGSVPGSI----- 1208
Db 418 AGTKCKGNLKGWMPKPSRNPRWGQAVFSGGRTVGSVANVTIRVAFATEKPKALIYSSIELV 477
QY 1209 --TKGIPSTRVP-----SDSAITYRGSITHTGTPADVLYKGTIT-----RIIGEDSPSR 1254
Db 478 VYNTGATLIRVPIAANVTIRSOIRCPGFLTYGTTPIAGYPTGIDATTNPNWKIAG-----VR 533
QY 1255 LDRGREDSLPK 1265
Db 534 INMGAGNKKPK 544
RESULT 92
PSC DROME
ID_PSC DROME STANDARD; PRT; 1603 AA.
AC P35820;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Posterior sex combs protein.
GN PSC.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92018190; PubMed=1833647;
RA Brunk B.P., Martin E.C., Sharp E., Adler P.N.;
RT "Drosophila genes Posterior Sex Combs and Suppressor two of zeste
encode proteins with homology to the murine bmi-1 oncogene";
RL Nature 353:351-353(1991).
CC -1- FUNCTION: The Polycomb group (Pc-G) genes are needed to maintain
expression patterns of the homeotic selector genes of the
antennapedia (Antp-C) and bithorax (Bx-C) complexes, and hence for
the maintenance of segmental determination.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Contains 1 RING-type zinc finger.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; X59275; CAA41965.1; -
PIR; S17983; S17983.
FlyBase; FBgn0005624; Psc.
InterPro; IPR001841; Znf_ring.
Pfam; PF00097; zf-C3HC4; 1.
SMART; SM00184; RING; 1.
DR PROSITE; PS00518; ZF_RING_1; 1.
DR PROSITE; PS00089; ZF_RING_2; 1.
KW Zinc-finger; Developmental protein; DNA-binding; Nuclear protein.
FT DOMAIN 47 53 POLY-THR.
FT DOMAIN 83 88 POLY-THR.
FT DOMAIN 91 98 POLY-THR.
FT DOMAIN 145 152 POLY-THR.
FT DOMAIN 184 202 POLY-SER.
FT ZN_FING 265 304 RING-TYPE.
FT DOMAIN 642 651 POLY-SER.
FT DOMAIN 1066 1069 POLY-GLY.
FT DOMAIN 1185 1189 POLY-PRO.


```

CC CC FUNCTION: Collagen type II is specific for cartilaginous tissues.
CC CC SUBUNIT: Homotrimers of alpha 1(II) chains.
CC CC ALTERNATIVE PRODUCTS:
CC CC Event=Alternative splicing; Named isoforms=2;
CC CC Name=Long;
CC CC IsoId=228481-1; Sequence=Displayed;
CC CC Name=Short;
CC CC IsoId=228481-2; Sequence=VSP_001139, VSP_001140;
CC CC PTM: Prolines at the third position of the tripeptide repeating
CC CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC CC SIMILARITY: Contains 1 VWF domain.
CC CC
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC
CC CC EMBL; M65161; AAA68100.1; -
CC CC EMBL; X57982; CAA41047.1; -
CC CC MGI; MGI:88452; Col2a1.
CC CC InterPro; IPR008161; C1g_helix.
CC CC InterPro; IPR008160; Collagen.
CC CC InterPro; IPR000885; Fib_collagen_C.
CC CC InterPro; IPR001007; VWF_C.
CC CC Pfam; PF01410; COLFI; 1.
CC CC Pfam; PF01391; Collagen; 18.
CC CC Pfam; PF00093; vwc; 1.
CC CC ProDom; PD000007; C1g_helix; 5.
CC CC ProDom; PD002078; Fib_collagen_C; 1.
CC CC SMART; SM00038; COLFI; 1.
CC CC SMART; SM00214; VWC; 1.
CC CC PROSITE; PS01208; VWF_C_1; 1.
CC CC PROSITE; PS0184; VWF_C_2; 1.
CC CC Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal; Alternative splicing.
FT SIGNAL 1 25
FT PROPEP 26 153
FT CHAIN 154 1213
FT PROPEP 1214 1459
FT DOMAIN 32 89
FT DOMAIN 173 1186
FT DOMAIN 1187 1213
FT VARSPIC 29 29
FT VARSPIC 30 98
FT VARSPIC 30 98
FT SEQUENCE 1459 AA; 139154 MW; F6C84FA7C532B7F2 CRC64;
Query Match 2.4%; Score 315.5; DB 1; Length 1459;
Best Local Similarity 21.2%; Pred. No. 0.0016;
Matches 353; Conservative 114; Mismatches 620; Indels 575; Gaps 79;
QY 711 SGN---EEEMVEAEALHASNEVPRGECSC--PATVNNSSDTEISPSHTEAAKDTQGN 765
DB 59 TGNVLDDIICEDPDCLNP---EIPFGECCPICADLATSCKLKGPKGQGEFGDIRDG 115
QY 766 GPKP-----PATLGADGP--PPGPTTPRRTS---RAPIEPTPASEATGA----- 805
DB 116 APGRGRDGRGTGPNPGAGPPGPPGPGSLAGNFAAQAGGYDEXAGGAGMVGQPM 175
QY 806 -PTPPAPPSPS-APPPVVPKEEETAAAPPVEGE-----BQKPPAAELAVDT 855
DB 176 GPMGPRGPPGAGAGP-----QGFQGNPGEPGEPGVSGPMGRGPP----- 217
QY 856 GKAEPEVKSECTEAEAGPAKGDAAEAATAEALKAEEKGSGRATTAKSSGAPQDS 915
DB 218 GPAGKP-----GDDGEA--GKPKSGERGLPDMGARGPPGTPGLPGVKHGKGYP--- 265
QY 916 DSSATCSADEYDEAGGDKNRLLSFRPFSLLTPTGDPANASPKPLDLKOLKORAAAIPTP 975

```

```

Db 266 -----GLDGAKGE-----AGAPGVKGESESGENSGP-GPMGPRGLPGERGRTP 309
QY 976 IQVTKVHEPPREDAAPTAKPAPPAPPPQNLQPSDAPQPGSSPRGKRSAPPADKEAF 1035
Db 310 AGAAGAR---GNDQPGPAGPPGPGVAGGPGFPGAPCAKCAKEAGPTGARGP----- 357
QY 1036 AAEAKQLPGDPPCWTSGLPFVPPPREVIKASPHADPSAFSAYAPPGHPLPLGLHDTARPV 1095
Db 358 -EGAQGSRGEP-----GNPSPGPAGASGNPCTDGI PKAGKSAGAPG 398
QY 1096 LPRPTTISNPPLISSAKHPSVLERQIGAISQGSVQLHVHYSEHAKAPVGPVTMGLPLP 1155
Db 399 IAGAPGPPGP-----RGPPGP--QGATGP 420
QY 1156 MDPKKLA---PFGSVKQEQSLSPRQAGPPESLGVPTAQEAASVLRGTALGSPVGGSIITKGI 1212
Db 421 LGPKQGAEGPIAGFKDQD-GPKGETGAGQAGPAGPAGEGKRG-ARGE-PGGAGPTGP 477
QY 1213 PSTR-VFSDSAITYRGSIT--HGTPADVLYKGTITRIIGEDSPSRL-----DRGR-- 1259
Db 478 PGERGAPGNRGFPQDGLAGPKGAP-----GERGPSGLAGPKANGDPGRPG 524
QY 1260 EDSLPGKHVIYEGKKHVLSEYEGMSVTQCSKEDGRSSSGPPHETAA-----PK 1308
Db 525 EPLGPGARGL-TGRPGDA-GPQGVKVPSPGAPGEDGR--PGPPPGQAGRGQGVMGFPQPK 580
QY 1309 RTYDMMEGRVGRA-----ISSASIEGLMGR-----AIPPERHSPHHLKEQHHRG-SIT 1356
Db 581 GA-----NGEPGKAGEKGLAGAGLUGLPGKDGCTGAAGPPGSPGAGERGQGAPGSGF 636
QY 1357 QGIPRSYVEAQEDYLRREAKLLKEGTTPPPPP---SRDLTEAYKTQALGLKLPKPAHEG 1413
Db 637 QGLP-----GPPPGEGGKQDQGIPEAGAP-----G 665
QY 1414 LVATVKEAGRSIHPIREE-----LHTEPLAPLPKEGSI-----TQGTTP 1456
Db 666 LVGPRGERG-----FPGERSPGAQGLQGRGLPTGTDTGP--KGAAGPDPGPAQAGPP 718
QY 1457 ----LKVDYTGASTTGSKKHVRSLIGSPGRTPFP-----VHPLDVMADALERACY 1504
Db 719 GLQMPGERGAAGIAGPKGD-RGDVGEKPEBAGPKDGGRLTGPIGPPGAGA----- 771
QY 1505 EESLKSPPGTASSSGSIARGAPVIVPELGPSPSLTYEDHGAPFAFAGHLPRGSPVTMRE 1564
Db 772 -NGEAGEAGPPSGSTGARGAPCEGETGP--GP-----AGFAG--PGAD---GQ 816
QY 1565 PTPRLQEGSSSSKASQDRKLTSTPREIAKSPHSTVPEHHPHIPISPYEHLHRLGVSVDLY 1624
Db 817 PGAKGDQ-----EAGQ-----KGDAGAPGQPSGAPGQGP-----TGVTPKGA 858
QY 1625 RSHIPLAFDPTSI PRGIPLDAAAAYLPRHLAPNPTYPHLYPPYLI RGYPDTALENROT 1684
Db 859 RG-----AQGP---PGATGFGAAGRVGPPGANGNP-----PAGPPGAGKDGPKG 902
QY 1685 IINDY-----ITSQMHNTATAMAQRA-----DMLRGLSPRESSLALYAAQPRGIIIDL-- 1734
Db 903 VRGDSGPPGRAGDPLGSPGAPAGEKGEPCGDDGSLDGGPPGQGL---AQORGIVGLPG 959
QY 1735 ----SQVPHL--PVLVPTTGTATAMDRLAYLPTAPQPFSSRHSPLSGPGTHLTKP 1789
Db 960 QRGERGFPGLPGPSGEKQKQAGPAGSGDR-----GPPGVGPPGLT-- 1000
QY 1789 TTTSSSERDRDRDRDREREKSIILSTTTVEHAPITWRPCTQSQSSSSSSGGGSSS 1848
Db 1001 -----GAGEGREGSGPAGDPPGRDGA-- 1024
QY 1849 SRPASHSHAHQHSPI SPRTDALQORPSVLHNTGMKGIITAVERSKPTVLRSTSTSSVR 1908
Db 1025 -----GVKG-----DRGETGALGA 1038
QY 1909 PATTFPPATHCPGL-----GTLDGYYTLMBEVLLPKAPRVARPERPRAD--- 1954

```

Db 1039 PGAPGPGSPGAGTCKQGDREAGAGQPMGSPS-----GPAGARGIAGPQPCGDKGE 1092
Qy 1955 -----TCHAFLEKPPARSGLEPPASSKSGSEPPPLVPPVPSGHATARTPA 1999
Db 1093 SGEGERGLKGRGFTGLQGLPGPPGSGDQSGAGPAGSPGR----- 1135
Qy 2000 KNLAPHASPDPPAPPASADPHREKTQSKPFSIQLELRSLGYSYSPSGVPSVP 2059
Db 1136 -----GPGCPVGPS-----GKGSNGIPGIPGPPGR 1162
Qy 2060 SPSLSLTKDKLPHLEELDLSHLEBELPKQPGVKLGGEAAHLPHRLPLPSQSSSPL 2119
Db 1163 GRSGETGVPVPG-----SPGPPGPPGPPGIDMSAFAGLGQREKGPDM 1208
Qy 2120 LQTAGVKGHORVVTLAHLSEVITDTRH--HPQOLSAPLPAPLYSPFGASCPVLDLR 2177
Db 1209 QY-----MRDEADSTLRQHDVEV---DATLKLNNQIESIRSPDGRSRKNPARTCDQLKLC 1261
Qy 2178 RP--PSDLYLPPDPHAGARGSP---HSEGGKRSPEPNKTSV 2214
Db 1262 HPEWKSQDYWDNQCTLDAMKVCNMTGETCVYVNPATV 1303

RESULT 94

CA25_HUMAN
ID CA25_HUMAN STANDARD; PRT; 1496 AA.
AC P05937;
DT 01-APR-1988 (Rel. 07, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 2(V) chain precursor.
GN COL5A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP SEQUENCE OF 1-463 FROM N.A.
RX MEDLINE=89123368; PubMed=2914927;
RA Woodbury D., Benson-Chanda V., Ramirez F.;
RT "Amino-terminal propeptide of human pro-alpha 2(V) collagen conforms
RT to the structural criteria of a fibrillar procollagen molecule.";
J. Biol. Chem. 264:2735-2738(1989).
RN [2]
RP SEQUENCE OF 398-1496 FROM N.A.
RX MEDLINE=87146331; PubMed=3029669;
RA Weil D., Bernard M.P., Gargano S., Ramirez F.;
RT "The pro alpha 2(V) collagen gene is evolutionarily related to the
RT major fibrillar-forming collagens";
Nucleic Acids Res. 15:181-198(1987).
RN [3]
RP SEQUENCE OF 1227-1496 FROM N.A.
RX MEDLINE=85289337; PubMed=2411731;
RA Myers J.C., Loidl H.R., Seyer J.M., Dion A.S.;
RT "Complete primary structure of the human alpha 2 type V procollagen
RT COOH-terminal propeptide";
J. Biol. Chem. 260:11216-11222(1985).
RN [4]
RP SEQUENCE OF 1449-1496 FROM N.A.
RX MEDLINE=89138450; PubMed=3224983;
RA Tsipouras P., Schwartz R.C., Liddell A.C., Salkeld C.S., Weil D.,
RA Ramirez F.;
RT "Genetic distance of two fibrillar collagen loci, COL3A1 and COL5A2,
RT located on the long arm of human chromosome 2";
Genomics 3:275-277(1988).
RN [5]
RP SEQUENCE OF 208-227.
RX MEDLINE=92239022; PubMed=1571108;
RA Mann K.;
RT "Isolation of the alpha 3-chain of human type V collagen and
RT characterization by partial sequencing";
Biol. Chem. Hoppe-Seyler 373:69-75(1992).

[6]
RP SEQUENCE OF 288-297 AND 606-617.
RC TISSUE=Bone;
RX MEDLINE=94237164; PubMed=8181482;
RA Moradi-Ameli M., Rousseau J.C., Klemm J.P., Champlaud M.F.,
RA Boutillon M.M., Bernillon J., Wallach J.M., van der Rest M.;
RT "Diversity in the processing events at the N-terminus of type-V
RT collagen";
Eur. J. Biochem. 221:987-995(1994).
RN [7]
RP DISEASE
RX MEDLINE=98087576; PubMed=9425231;
RA Michalickova K., Susic M., Willing M.C., Wenstrup R.J., Cole W.G.;
RT "Mutations of the alpha2(V) chain of type V collagen impair matrix
RT assembly and produce Ehlers-Danlos syndrome type I";
Hum. Mol. Genet. 7:249-255(1998).
RN [8]
RP VARIANT EDS-II ARG-960.
RX MEDLINE=98455031; PubMed=9783710;
RA Richards A.J., Martin S., Nicholls A.C., Harrison J.B., Pope F.M.,
RA Burrows N.P.;
RT "A single base mutation in COL5A2 causes Ehlers-Danlos syndrome type
RT II";
J. Med. Genet. 35:846-848(1998).
CC -1- FUNCTION: Type V collagen is a member of group I collagen
CC (fibrillar forming collagen). It is a minor connective tissue
CC component of nearly ubiquitous distribution. Type V collagen binds
CC to DNA, heparan sulfate, thrombospondin, heparin, and insulin.
CC -1- SUBUNIT: Trimers of two alpha 1(V) and one alpha 2(V) chains in
CC most tissues and trimers of one alpha 1(V), one alpha 2(V), and
CC one alpha 3(V) chains in placenta.
CC -1- PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC -1- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC type I (EDS-I) [MIM:130000]; also known as Ehlers-Danlos syndrome
CC gravis. EDS-I is an autosomal dominant connective-tissue disorder
CC characterized by loose-jointedness and fragile, velvety,
CC stretchable, bruisable skin that heals with peculiar 'cigarette-
CC paper' scars. Inheritance is autosomal dominant.
CC -1- DISEASE: Defects in COL5A2 are a cause of Ehlers-Danlos syndrome
CC type II (EDS-II) [MIM:130010]; also known as Ehlers-Danlos
CC syndrome mitis. Inheritance is autosomal dominant.
CC -1- SIMILARITY: Contains 1 VWFC domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J04478; AA51859.1; -
DR EMBL; X04758; CAA28454.1; -
DR EMBL; M11718; AAA52058.1; -
DR PIR; A31427; GCHU2V.
DR PDB; 1A9A; 18-NOV-98.
DR Genew; HGNC:2210; COL5A2.
DR MIM; 120190; -
DR MIM; 130000; -
DR MIM; 130010; -
DR GO; GO:0005588; C:collagen type V; TAS.
DR GO; GO:0008151; P:cell growth and/or maintenance; TAS.
DR InterPro; IPR008161; C1g_helix.
DR InterPro; IPR008160; Collagen.
DR InterPro; IPR000885; Fib_collagen_C.
DR InterPro; IPR002181; Fibrinogen_C.
DR Pfam; PF01410; COLFI; 1.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF00093; vwc; 1.
DR ProDom; PD000007; C1g_helix; 5; 1.
DR ProDom; PD002078; Fib_collagen_C; 1.

DT 10-OCT-2003 (Rel. 42, Created)
 DT 10-OCT-2003 (Rel. 42, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Muscle M-line assembly protein unc-89 (Uncoordinated protein 89).
 GN UNC-89 OR C09D1.1.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A., FUNCTION, AND TISSUE SPECIFICITY.
 RC STRAIN=Bristol N2;
 RX MEDLINE=96180278; PubMed=8603916;
 RA Benian G.M., Tinley T.L., Tang X., Borodovsky M.;
 RT "The Caenorhabditis elegans gene unc-89, required for muscle M-line
 assembly, encodes a giant modular protein composed of Ig and signal
 transduction domains.";
 RT J. Cell Biol. 132:835-848(1996).
 RL [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=Bristol N2;
 RC Du Z., Le T.T., Wilson R.;
 RA Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RL [3]
 RP REVISIONS.
 RA Waterston R.;
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: Structural component of the muscle M-line. Myofibril
 CC lattice assembly begins with positional cues laid down in the
 CC basement membrane and muscle cell membrane. UNC-89 responds to
 CC these signals, localizes, and then participates in assembling an
 CC M-line.
 CC -!- TISSUE SPECIFICITY: Localizes to the middle of A-bands.
 CC -!- SIMILARITY: Contains 1 DBL-homology (DH) domain.
 CC -!- SIMILARITY: Contains 1 fibronectin type III domain.
 CC -!- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
 CC -!- SIMILARITY: Contains 1 PH domain.
 CC -!- SIMILARITY: Contains 5 RCSD domains.
 CC -!- SIMILARITY: Contains 1 SH3 domain.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U33058; AAB00542.1; -;
 DR EMBL; AF003131; AAB54132.2; -;
 DR PDB; 1FHO; 20-DEC-00.
 DR WormPep; C09D1.1; CE30426.
 DR InterPro; IPR008957; FN III-like.
 DR InterPro; IPR003961; FN III.
 DR InterPro; IPR007110; Ig-like.
 DR InterPro; IPR003598; Ig_C2.
 DR InterPro; IPR003006; Ig_MHC.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR007850; RCSD.
 DR InterPro; IPR000219; RhoGEF.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00041; fn3; 1.
 DR Pfam; PF00047; ig; 47.
 DR Pfam; PF00165; fh; 1.
 DR Pfam; PF05177; RCSD; 5.
 DR Pfam; PF00621; RhoGEF; 1.
 DR Pfam; PF00018; SH3; 1.
 DR SMART; SM00408; IGC2; 23.
 DR SMART; SM00325; RhoGEF; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PSS0010; DH_2; 1.
 DR PROSITE; PSS0835; IG LIKE; 49.
 DR PROSITE; PSS0003; PH_DOMAIN; 1.
 DR PROSITE; PSS0002; SH3; 1.
 DR Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
 DR 3D-structure.
 KW FT DOMAIN 63 127 SH3.
 FT DOMAIN 152 330 DH.
 FT DOMAIN 342 498 PH.
 FT DOMAIN 547 633 IG-LIKE C2-TYPE 1.
 FT DOMAIN 648 736 IG-LIKE C2-TYPE 2.
 FT DOMAIN 748 838 IG-LIKE C2-TYPE 3.
 FT DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
 FT DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
 FT DOMAIN 1140 1227 IG-LIKE C2-TYPE 6.
 FT DOMAIN 1272 1315 THR-RICH.
 FT DOMAIN 1375 1475 RCSD 1.
 FT DOMAIN 1479 1585 RCSD 2.
 FT DOMAIN 1597 1695 RCSD 3.
 FT DOMAIN 1700 1799 RCSD 4.
 FT DOMAIN 1800 1860 RCSD 5.
 FT DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
 FT DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
 FT DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
 FT DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
 FT DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
 FT DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
 FT DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
 FT DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
 FT DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
 FT DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
 FT DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
 FT DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
 FT DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
 FT DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
 FT DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.
 FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
 FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
 FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
 FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
 FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
 FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
 FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
 FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
 FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
 FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
 FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
 FT DOMAIN 4588 4678 IG-LIKE C2-TYPE 33.
 FT DOMAIN 4681 4771 IG-LIKE C2-TYPE 34.
 FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
 FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
 FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
 FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
 FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
 FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
 FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
 FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
 FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
 FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
 FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
 FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
 FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
 FT DOMAIN 6275 6368 FIBRONECTIN TYPE-III.
 FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
 FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
 FT DISULFID 568 621 POTENTIAL.
 FT DISULFID 2908 2975 POTENTIAL.
 FT DISULFID 3015 3065 POTENTIAL.
 FT DISULFID 3707 3759 POTENTIAL.
 FT DISULFID 3826 3890 POTENTIAL.
 FT DISULFID 5092 5157 POTENTIAL.
 FT DISULFID 5298 5350 POTENTIAL.
 FT DISULFID 5508 5560 POTENTIAL.
 FT DISULFID 5616 5669 POTENTIAL.
 FT DISULFID 5722 5764 POTENTIAL.
 FT DISULFID 5836 5901 POTENTIAL.

RESULT 96
MUC1 HUMAN STANDARD: PRT: 1255 AA.
AC P15941; P15942; P17626; Q14128; Q14876; Q16437; Q16442;
AC Q16615; Q9BXA4; Q9UE75; Q9UE76; Q9UQL1; Q9Y4U2;
DT 01-JAN-1990 (Rel. 13, Created)
DT 10-OCT-2003 (Rel. 14, Last sequence update)
DE Mucin 1 precursor (MUC-1) (Polymorphic epithelial mucin) (PEM) (PEMT)
DE (Episialin) (Tumor-associated mucin) (Carcinoma-associated mucin)
DE (Tumor-associated epithelial membrane antigen) (EMA) (H23AG) (Peanut-
DE reactive urinary mucin) (PUM) (Breast carcinoma-associated antigen
DE DF3) (CD227 antigen).
GN MUC1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Pancreas;
RX MEDLINE=90368716; PubMed=2394722;
RA Lan M.S., Batra S.K., Qi W.-N., Metzgar R.S., Hollingsworth M.A.;
RT "Cloning and sequencing of a human pancreatic tumor mucin cDNA.";
RL J. Biol. Chem. 265:15294-15299(1990).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE=90202794; PubMed=2318825;
RA Lightenberg M.J.L., Vos H.L., Gennissen A.M.C., Hilkens J.;
RT "Episialin, a carcinoma-associated mucin, is generated by a
RT polymorphic gene encoding splice variants with alternative amino
RT termini.";
RL J. Biol. Chem. 265:5573-5578(1990).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90368715; PubMed=1697589;
RA Gendler S.J., Lancaster C.A., Taylor-Papadimitriou J., Duhig T.,
RA Feat N., Burchell J., Pemberton L., Lalani E.-N., Wilson D.;
RT "Molecular cloning and expression of the human tumor-associated
RT polymorphic epithelial mucin.";
RL J. Biol. Chem. 265:15286-15293(1990).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=91097524; PubMed=2268309;
RA Lancaster C.A., Feat N., Duhig T., Wilson D.,
RA Taylor-Papadimitriou J., Gendler S.J.;
RT "Structure and expression of the human polymorphic epithelial mucin
RT gene: an expressed VNTR unit.";
RL Biochem. Biophys. Res. Commun. 173:1019-1029(1990).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 5).
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276413; PubMed=2351132;
RA Wreschner D.H., Hareuveni M., Tsarfaty I., Smorodinsky N., Horev J.,
RA Zaretsky J., Kotkes P., Weiss M., Lathe R., Dion A., Keydar I.;
RT "Human epithelial tumor antigen cDNA sequences. Differential splicing
RT may generate multiple protein forms.";
RL Eur. J. Biochem. 189:463-473(1990).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Breast carcinoma;
RX MEDLINE=90276414; PubMed=2112460;
RA Hareuveni M., Tsarfaty I., Zaretsky J., Kotkes P., Horev J.,
RA Zrihan S., Weiss M., Green S., Lathe R., Keydar I., Wreschner D.H.;
RT "A transcribed gene, containing a variable number of tandem repeats,
RT codes for a human epithelial tumor antigen. cDNA cloning, expression
RT of the transcribed gene and over-expression in breast cancer
RT tissue.";
RL Eur. J. Biochem. 189:475-486(1990).
RN [7]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=91033045; PubMed=1688329;

RA Tsarfaty I., Hareuveni M., Horev J., Zaretsky J., Weiss M.,
RA Jeltsch J.M., Garnier J.M., Lathe R., Keydar I., Wreschner D.H.;
RT "Isolation and characterization of an expressed hypervariable gene
RT coding for a breast-cancer-associated antigen.";
RN Gen 93:313-318(1990).
RN [8]
RP SEQUENCE FROM N.A. (ISOFORM 7).
RX MEDLINE=95010060; PubMed=7925397;
RA Zrihan-Licht S., Vos H.L., Baruch A., Elroy-Stein O., Sagiv D.,
RA Keydar I., Hilkens J., Wreschner D.H.;
RT "Characterization and molecular cloning of a novel MUC1 protein,
RT devoid of tandem repeats, expressed in human breast cancer tissue.";
RL Eur. J. Biochem. 224:787-795(1994).
RN [9]
RP SEQUENCE FROM N.A. (ISOFORMS 6; 7 AND 8).
RX MEDLINE=97355747; PubMed=9212228;
RA Oosterkamp H.M., Scheiner L., Stefanova M.C., Lloyd K.O.,
RA Finstad C.L.;
RT "Comparison of MUC-1 mucin expression in epithelial and non-epithelial
RT cancer cell lines and demonstration of a new short variant form
RT (MUC-1/Z).";
RL Int. J. Cancer 72:87-94(1997).
RN [10]
RP SEQUENCE FROM N.A. (ISOFORM 7).
RX Zhang L.X., Li C.H.;
RT "Molecular cloning of an isoform of MUC1, MUC1/Y.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [11]
RP SEQUENCE FROM N.A. (ISOFORM 9).
RC TISSUE=Epithelial cancer;
RX Zhang L.X., Li C.H., Sun L.Y., Yue W.;
RT "Cloning of a new potential secreted short variant form of MUC1 mucin
RT in epithelial cancer cell line.";
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [12]
RP PARTIAL SEQUENCE FROM N.A.
RX MEDLINE=88330762; PubMed=3417635;
RA Gendler S.J., Taylor-Papadimitriou J., Duhig T., Rothbard J.,
RA Burchell J.;
RT "A highly immunogenic region of a human polymorphic epithelial mucin
RT expressed by carcinomas is made up of tandem repeats.";
RL J. Biol. Chem. 263:12820-12823(1988).
RN [13]
RP SEQUENCE OF 1-160 FROM N.A. (ISOFORM 2).
RX MEDLINE=90088473; PubMed=2597151;
RA Abe M., Siddiqui J., Kufe D.;
RT "Sequence analysis of the 5' region of the human DF3 breast
RT carcinoma-associated antigen gene.";
RL Biochem. Biophys. Res. Commun. 165:644-649(1989).
RN [14]
RP SEQUENCE OF 1-109 FROM N.A. (ISOFORM 2).
RC TISSUE=Thyroid;
RX MEDLINE=96183746; PubMed=8608966;
RA Weiss M., Baruch A., Keydar I., Wreschner D.H.;
RT "Preoperative diagnosis of thyroid papillary carcinoma by reverse
RT transcriptase polymerase chain reaction of the MUC1 gene.";
RL Int. J. Cancer 66:55-59(1996).
RN [15]
RP SEQUENCE OF 1-89 FROM N.A.
RC TISSUE=Lung;
RX MEDLINE=96181716; PubMed=8604237;
RA Yu C.J., Yang P.C., Shew J.Y., Hong T.M., Yang S.C., Lee Y.C.,
RA Lee L.N., Luh K.T., Wu C.W.;
RT "Mucin mRNA expression in lung adenocarcinoma cell lines and
RT tissues.";
RL Oncology 53:118-126(1996).
RN [16]
RP SEQUENCE OF 1-46 FROM N.A. (ISOFORMS 3 AND 4).
RC TISSUE=Breast carcinoma;
RX Buluwela L., Liu Q., Luqmani Y.A., Gomm J.J., Coombes R.C.;
RL Submitted (OCT-1992) to the EMBL/GenBank/DBJ databases.
RN [17]
RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.

RX MEDLINE=97460054; PubMed=9312074;
 RA Mueller S., Goletz S., Packer N., Gooley A.A., Lawson A.M.,
 RA Hanisch F.-G.;
 RT "Localization of O-glycosylation sites on glycopeptide fragments from
 RT lactation-associated MUC1. All putative sites within the tandem
 RT repeat are glycosylation targets in vivo.";
 RL J. Biol. Chem. 272:24780-24793(1997).
 RN [18]
 RP CARBOHYDRATE-LINKAGE SITES WITHIN THE REPEAT.
 RX MEDLINE=99303572; PubMed=10373415;
 RA Mueller S., Alving K., Peter-Katalinic J., Zachara N., Gooley A.A.,
 RA Hanisch F.-G.;
 RT "High density O-glycosylation on tandem repeat peptide from secretory
 RT MUC1 of T47D breast cancer cells.";
 RL J. Biol. Chem. 274:18165-18172(1999).
 RN [19]
 RP POLYMORPHISM WITHIN THE REPEAT.
 RX MEDLINE=21359366; PubMed=11350974;
 RA Engelmann K., Baldus S.E., Hanisch F.-G.;
 RT "Identification and topology of variant sequences within individual
 RT repeat domains of the human epithelial tumor mucin MUC1.";
 RL J. Biol. Chem. 276:27764-27769(2001).
 RN [20]
 RP CHARACTERIZATION OF ISOFORM Y, AND MUTAGENESIS OF ASP-1116.
 RX MEDLINE=99211485; PubMed=10197628;
 RA Baruch A., Hartmann M.-L., Yoeli M., Adereh Y., Greenstein S.,
 RA Stadler Y., Skornik Y., Zaretsky J., Smorodinsky N.I., Keydar I.,
 RA Wreschner D.H.;
 RT "The breast cancer-associated MUC1 gene generates both a receptor and
 RT its cognate binding protein.";
 RL Cancer Res. 59:1552-1561(1999).
 RN [21]
 RP PARTIAL SEQUENCE, AND CHARACTERIZATION OF CLEAVAGE SITE.
 RX MEDLINE=21240104; PubMed=11341784;
 RA Parry S., Silverman H.S., McDermott K., Willis A., Hollingsworth M.A.,
 RA Harris A.;
 RT "Identification of MUC1 proteolytic cleavage sites in vivo.";
 RL Biochem. Biophys. Res. Commun. 283:715-720(2001).
 RN [22]
 RP CHARACTERIZATION.
 RX MEDLINE=21836452; PubMed=11847293;
 RA Wreschner D.H., McGuckin M.A., Williams S.J., Baruch A., Yoeli M.,
 RA Ziv R., Okun L., Zaretsky J., Smorodinsky N., Keydar I., Neophytou P.,
 RA Stacey M., Lin H.-H., Gordon S.;
 RT "Generation of ligand-receptor alliances by 'SEA' module-mediated
 RT cleavage of membrane-associated mucin proteins.";
 RL Protein Sci. 11:698-706(2002).
 RN [23]
 RP PHOSPHORYLATION.
 RX MEDLINE=95080414; PubMed=7988707;
 RA Zrihan-Licht S., Baruch A., Elroy-Stein O., Keydar I., Wreschner D.H.;
 RT "Tyrosine phosphorylation of the MUC1 breast cancer membrane proteins.
 RT Cytokine receptor-like molecules.";
 RL FEBS Lett. 356:130-136(1994).
 CC -1- FUNCTION: May play a role in adhesive functions and in cell-cell
 CC interactions, metastasis and signaling. May provide a protective

Query Match 2.4%; Score 314; DB 1; Length 1255;
 Best Local Similarity 21.0%; Pred. No. 0.0016;
 Matches 322; Conservative 114; Mismatches 548; Indels 546; Gaps 75;
 725 HASGNEVPRGCSGPATVNS--SDTE-----SIPSPHTEAAKDTGQNGKPPA 771
 26 HAS--STPGKEKETSATORSVPSTENKAVMSVSSVLSHSPGSGSSTTQGDVTLAPA 83
 772 TLGADG-----PPGPPTPPRR--TSRAPTEPTPASEA-----TGAPT 807
 84 TEPASGSAATWGQDVTSVPVTRPALGSTTPPAHDVTSAPDNKPAFGSTAPPAGHGVTSAPD 143
 808 PPRAPPSAPPVVPVPEKEEETAAAPPVEGEQKPPAAEELAVDTGKA---EPPVKS 864
 144 TRPAPGSTAPPAGHGV---TSAPDTRPAP-----GSTAPPAGHGVTSAPDTRPAPGSTAPP 196

QY 865 ECTEEAEEGPAGKADAEAAEATAEALKAEKKEGSGRATTAKSGAPQDDSDSATCSAD 924
 DB 197 GVTSAEDTRPAGSTAPPA-----HGVTSAEDTRPAGSTAPPAHGVTSAED 243
 QY 925 EVDEAEGDKNRLSPSPSLTPTGDPANASPKPLDLKQKORAAAIPIQ-VTKVHE 983
 DB 244 -----TRPAGSTAPPAGHGVTSAPDTRPAP-----GSTAPPAGHGV 279
 QY 984 PPREDAAFTKAP-PAPPPPPONQOPESDAPQOPQSPGKSRSPAPPADKAEFAAEAKL 1042
 DB 280 ----SAPDTRPAGSTAPPAGHGVTSAPDTRPAGST-----APPAHGVTSAEDTRPA 327
 QY 1043 PGDPCCWTSGLFPVPPREVIKASPHAPDPSAFYAPPGHPLPLGLDHTDARVLPRTPTI 1102
 DB 328 PGS-----TAPPAGHGVTSAPDT-RPAPGSTAPPAGHGV-----TSAPDTRPAGST 371
 QY 1103 SNPPPLISSAKHPSVLEROICAIQOGMSVOLHVPSYSEHAKAPVGVMTMGLPLMDPKLA 1162
 DB 372 APPAGHVTSA-----PDTRPAGSTAPPAGHGVTSAPDTRP-A 407
 QY 1163 PFGVGKQELSPRGQAGPPESLGVPTAQEASVLRGTALGSPGGSITKGIPTSTR-VPSDS 1221
 DB 408 P-----GSTAPP-AHGVTSAEDTRPAGST--APPAHGVTSA-PDTRPAGST 451
 QY 1222 AITVRSITHGTADVLYKGTITRIIGEDSPSLDRGREDSLPKGHVYEGKKGHVLSYE 1281
 DB 452 A-----PPAGHGVTSAPDTR-----PAPG-----STAPPAGHGV----- 478
 QY 1282 GGMSTQCKEDGRSSGPPHETAAPKRTYDMGVRGVRGAISSASIEGLMGRAIPPERHS 1341
 DB 479 -----TSAPDTRPA-----PGSTAPP-----AHGVTSA-----PDTRPA 507
 QY 1342 PHLKEQHIRGSIITQIGPRSYVEAQEDYLRRKALKREGTTPPPP-----PPSRDLTBEAY 1397
 DB 508 P-----GSTAP-PAGHVTSAED-----TRPAGSTAPPAGHVTSAED 542
 QY 1398 KTOALGPLKPAHEGLVATVKEAGRSIHEIPRELRHTPELPLAPRLKEGSIITQGTPL 1457
 DB 543 DTRPAGSTAPPAGH-GVTS-----APDTRPAP----- 568
 QY 1458 KYDTGASTTSGKKHIDVRSLLIG---SPGRTFPPVPLDVMADARALERACRYESLSKSRGT 1514
 DB 569 -----GSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTR-----PAPGS 610
 QY 1515 ASSSGGSIAGAPVIVPELGKPRQSLTYEDHGFAGHLPGRGSPVMTREPTPRLOEGLS 1574
 DB 611 TAPPAGHVT-SAPDTRPAGST-----APPA-HGVTSAEDT--RPAP----- 648
 QY 1575 SSSKASQDRKLTSTPREIAKSPHSTVPPEHHPHPISPYEHLRGSVDLYRSHIPLAPDP 1634
 DB 649 -GSTAPPAGHGVTSAP-DTRPAGSTAPPAGHGVTSAPDTRPAG-----STAPPAGHV 698
 QY 1635 TSIPRGIPLDAAAYYLPRHLAPNPTYPHLYPPYLYIRGYPTDPALENRQTIINDYITQQ 1694
 DB 699 TSAPDTRP-----APGSTAP---PAHGVTSAEDTRPAGSTAPPAGHGVTSAP 742
 QY 1695 MHNTATMAQADMLRGLSPRESSLALNVAAGPRGIIDLSQVPHLPVLVPTPTGTPATA 1754
 DB 743 ---DTRPAGSTAPPAGHGVTSAPDT-----RPAGSTAPP 774
 QY 1755 MDRLAYLF-TAPQPFSSRRHSSPLSPGQPTH--LTKPTTTSSSRERDRDRDRERE 1811
 DB 775 AHGVTSAEDTRPAP-----GSTAP-----PAHGVTSAEDTRPAP----- 808
 QY 1812 KSILTSITTTVEHAPIWRPTEQSSGSGSGSGSGSGSS--SRPASHSHAHQHSPTSPTQD 1869
 DB 809 ----GSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPAGSTA---PPAHGVTS 860
 QY 1870 ALQORPSVLHNTKMGIIITAVEPSKPTVLRSTSSSPVRPAATPPATHCPLGGLDGVY 1929
 DB 861 APDTRPAGSTAPPAGHGVTSAPDTRPA-----PGSTAPPA-----HGV- 898
 QY 1930 PTLMEVLLPKPEAPRVARPERPRADTGHAFKAPPAR-----SGLEPASSPSKSGSEPRPLV 1985

```

Db      899  -----TSAPDTRPAGSTAPPAGHVTSPADTRPAGSTAPPAGHVTSPADNRPAL 948
QY      1986  -----PPV-----SCHATIARTPAKNAAPHASDPAPPASASDPHREKTSQKPSIOE 2035
Db      949  GSTAPPVHNVTASGSAS-----GSASTLVHNGTSARATTPAS-----KSTPFSIP- 995
QY      2036  LELRSLGYHGSSYSPEGVSPVSPSLTHDKGLPKHLEELD KSHLEGELAPKQPGPVK 2095
Db      996  -----SHH-----SDTPTLASHSTKTDASSTHSSVP-----1023
QY      2096  LGGEAAHLPHLRPLPSPQSSSPLLQATPCVKGQRVWTLAQHISEVITQDYTRHHPQOL 2155
Db      1024  -----BLTSSNHSSTQSLST--GVS-----FFFLSFHISNL-----QF 1054
QY      2156  SAPLPAPLYSPFCGSCFVLDLRPPSPDLVL 2185
Db      1055  NSSLEDPTDY-----YQELQDISEMPL 1078

RESULT 97
CA24 HUMAN
ID CA24 HUMAN STANDARD; PRT; 1712 AA.
AC P08572;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 2(IV) chain precursor.
GN COL4A2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=89066769; PubMed=3198637;
RA Hostikka S.L., Tryggvason K.;
RT "The complete primary structure of the alpha 2 chain of human type IV
RT collagen and comparison with the alpha 1(IV) chain.";
RL J. Biol. Chem. 263:19488-19493(1988).
RN [2]
SEQUENCE OF 1-1042 FROM N.A.
RX MEDLINE=87219158; PubMed=3582677;
RA Hostikka S.L., Kurkinen M., Tryggvason K.;
RT "Nucleotide sequence coding for the human type IV collagen alpha 2
RT chain cDNA reveals extensive homology with the NC-1 domain of alpha 1
RT (IV) but not with the collagenous domain or 3'-untranslated region.";
RL FEBS Lett. 216:281-286(1987).
RN [4]
SEQUENCE OF 1451-1485 FROM N.A.
RX MEDLINE=87092438; PubMed=3025878;
RA Griffin C.A., Emanuel B.S., Hansen J.R., Cavenee W.K., Myers J.C.;
RT "Human collagen genes encoding basement membrane alpha 1(IV) and
RT alpha 2(IV) chains map to the distal long arm of chromosome 13.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:512-516(1987).
RN [5]
SEQUENCE OF 1486-1712 FROM N.A.
RX MEDLINE=87250571; PubMed=2439508;
RA Myers J.C., Howard P.S., Jelen A.M., Dion A.S., Macarak E.J.;
RT "Duplication of type IV collagen COOH-terminal repeats and species-
RT specific expression of alpha 1(IV) and alpha 2(IV) collagen genes.";
RL J. Biol. Chem. 262:9231-9238(1987).
RN [6]
SEQUENCE OF 1-33 FROM N.A.

```

```

RX MEDLINE=89034231; PubMed=3182844;
RA Soinen R., Huotari M., Hostikka S.L., Prockop D.J., Tryggvason K.;
RT "The structural genes for alpha 1 and alpha 2 chains of human type IV
RT collagen are divergently encoded on opposite DNA strands and have an
RT overlapping promoter region.";
RL J. Biol. Chem. 263:17217-17220(1988).
RN [7]
SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=89030632; PubMed=2846280;
RA Poeschl E., Pollner R., Kuehn K.;
RT "The genes for the alpha 1(IV) and alpha 2(IV) chains of human
RT basement membrane collagen type IV are arranged head-to-head and
RT separated by a bidirectional promoter of unique structure.";
RL EMBO J. 7:2687-2695(1988).
RN [8]
SEQUENCE OF 1-33 FROM N.A.
RX MEDLINE=93305049; PubMed=8317999;
RA Fischer G., Schmidt C., Opitz J., Cully Z., Kuehn K., Poeschl E.;
RT "Identification of a novel sequence element in the common promoter
RT region of human collagen type IV genes, involved in the regulation of
RT divergent transcription.";
RL Biochem. J. 292:687-695(1993).
RN [9]
SEQUENCE OF 1480-1535; 1545-1614; 1617-1701 AND 1705-1712.
RX MEDLINE=89005112; PubMed=2844531;
RA Siebold B., Deutzmann R., Kuehn K.;
RT "The arrangement of intra- and intermolecular disulfide bonds in the
RT carboxyterminal, non-collagenous aggregation and cross-linking domain
RT of basement-membrane type IV collagen.";
RL Eur. J. Biochem. 176:617-624(1988).
RN [10]
FUNCTION: Type IV collagen is the major structural component of
CC glomerular basement membranes (GBM), forming a 'chicken-wire'
CC meshwork together with laminins, proteoglycans and entactin/
CC nidogen.
CC [11]
SUBUNIT: There are six type IV collagen isoforms, alpha 1(IV)-
CC alpha 6(IV), each of which can form a triple helix structure
CC with 2 other chains to generate type IV collagen network.
CC [12]
DOMAIN: Alpha chains of type IV collagen have a noncollagenous
CC domain (NC1) at their C-terminus, frequent interruptions of the
CC G-X-Y repeats in the long central triple-helical domain (which may
CC cause flexibility in the triple helix), and a short N-terminal
CC triple-helical 7S domain.
CC [13]
PTM: Prolines at the third position of the tripeptide repeating
CC unit (G-X-Y) are hydroxylated in some or all of the chains.
CC [14]
PTM: Type IV collagens contain numerous cysteine residues which
CC are involved in inter- and intramolecular disulfide bonding. 12 of
CC these, located in the NC1 domain, are conserved in all known type
CC IV collagens.
CC [15]
This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC [16]
EMBL; X05562; CAA29076.1; -
DR EMBL; X05610; CAA29098.1; -
DR EMBL; J02760; AAA58422.1; -
DR EMBL; M36963; AAA53099.1; -
DR EMBL; X12784; CAA31275.1; -
DR EMBL; J04217; AAA53097.1; -
DR PIR; A32024; CGHU2B.
DR Genew; HGNC:2203; COL4A2.
DR MIM; 120090; -
DR GO; GO:0005587; C:collagen type IV; TAS.
DR GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
DR GO; GO:0030198; P:extracellular matrix organization and bioge. .; NAS.
DR InterPro; IPR008161; Clg helix.
DR InterPro; IPR008160; Collagen.

```


Qy 2366 NAFNPLNA---SASLPAAMPITRADGRSDHTLSPGGG-----GKAKVSG----- 2407
Db 2402 STFPVIOAGPQMLTIPAVSVIHRTVTSVGTITTEASGSPNRPTGVAEELSSWPCIPIGOI 2461
Qy 2408 -RPSRRKAKSPA-----PGLA-----S 2423
Db 2462 HVPGLONLSPPALQSLTSLGMEITVNLVGLANATVPGQHPHGLNALVAGLOVLNAPAQ 2521
Qy 2424 GDRPPSVSVHSGDCNRRPTPLNRVWEDRPPSAGSTPPPY-----NPLIMRLQ--AGV-M 2476
Db 2522 SPAPPA-----HIQGLQLNLALTPLTPSVGPVAVGTGTPTTAPNSKAMELQMPAQGH 2577
Qy 2477 ASPPPPGLPAG-----SGPLAGPHHAWDEE 2501
Db 2578 SAEPPOQSPGPOETPQTVSGPSA--DHARPED 2608

RESULT 99
PGCV MOUSE
ID PGCV MOUSE STANDARD; PRT; 3358 AA.
AC Q62059; Q62058; Q9CUT0;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Versican core protein precursor (Large fibroblast proteoglycan)
DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M).
GN CSPG2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS V0; V1 AND V2).
RC STRAIN=C57BL/6; and Swiss Webster; TISSUE=Brain;
RX MEDLINE=95122551; PubMed=7822336;
RA Ito K., Shinomura T., Zako M., Ujita M., Kimata K.;
RT "Multiple forms of mouse PG-M, a large chondroitin sulfate
RT proteoglycan generated by alternative splicing.";
RL J. Biol. Chem. 270:958-965(1995).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM V3).
RC STRAIN=C57BL/6;
RX MEDLINE=95181355; PubMed=7876137;
RA Zako M., Shinomura T., Ujita M., Ito K., Kimata K.;
RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
RT without a chondroitin sulfate attachment in region in mouse and human
RT tissues.";
RL J. Biol. Chem. 270:3914-3918(1995).
RN [3]
RP SEQUENCE OF 1-1692 FROM N.A. (ISOFORM V1).
RC STRAIN=C57BL/6J; TISSUE=Skin;
RX MEDLINE=23354683; PubMed=12466851;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yananaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Mateuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grammond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lennah R., Lyons P.A.,
RA Maglott D.R., Maitais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramchandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Walstead C., Wang Y., Watanabe Y., Wells C.,
RA Wilmig L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,

RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP INTERACTION WITH FBLN1
RX MEDLINE=93329059; PubMed=10400671;
RA Asberg A., Adam S., Kostka G., Timpl R., Heinigaard D.;
RT "Fibulin-1 is a ligand for the C-type lectin domains of aggrecan and
RT versican.";
RL J. Biol. Chem. 274:20444-20449(1999).
CC -!- FUNCTION: May play a role in intercellular signaling and in
CC connecting cells with the extracellular matrix. May take part in
CC the regulation of cell motility, growth and differentiation. Binds
CC hyaluronic acid.
CC -!- SUBUNIT: Interacts with FBLN1.
CC -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=4;
CC Comment-Additional isoforms seem to exist;
CC Name=V0;
CC IsoId=Q62059-1; Sequence=Displayed;
CC Name=V1;
CC IsoId=Q62059-2; Sequence=VSP_003087, VSP_003088;
CC Name=V2;
CC IsoId=Q62059-3; Sequence=VSP_003089;
CC Name=V3;
CC IsoId=Q62059-4; Sequence=VSP_003087, VSP_003090;
CC -!- TISSUE SPECIFICITY: V2 is found only in brain.
CC -!- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -!- SIMILARITY: Contains 2 link domains.
CC -!- SIMILARITY: Contains 2 EGF-like domains.
CC -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -!- SIMILARITY: Belongs to the aggrecan/versican proteoglycan family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D16263; BAA03796.1; -;
CC EMBL; D28599; -; NOT_ANNOTATED_CDS.
CC EMBL; D32040; BAA06802.1; -;
CC EMBL; AK014525; BAB29411.2; -;
CC HSSP; P01132; 1EPG.
CC MGD; MGI:102889; Cspg2.
CC InterPro; IPR000152; Asx_hydroxyl_S.
CC InterPro; IPR000742; EGF_2.
CC InterPro; IPR001881; EGF_Ca.
CC InterPro; IPR006209; EGF-like.
CC InterPro; IPR007110; Ig-like.
CC InterPro; IPR003599; Ig.
CC InterPro; IPR001304; Lectin_C.
CC InterPro; IPR000538; Link.
CC InterPro; IPR000436; Sushi_SCR_CCP.
CC Pfam; PF00008; EGF; 2.
CC Pfam; PF00047; Ig; 1.
CC Pfam; PF00059; Lectin_c; 1.
CC Pfam; PF00084; sushi; 1.
CC Pfam; PF00193; Xlink; 2.
CC PRINTS; PD01265; LINKMODULE.
CC ProDom; PD000918; Link; 2.
CC SMART; SM00032; CCP; 1.

QY 951 PRANASQKPLDLKQLKORAAAIPIQVTKVHPPREDAAPTAPAPPPQN---LQ 1006
Db 1389 QLVTVTPKDP-----EAAERGOYEVAPSNQFDDSATDTHQFILAE 1432
QY 1007 PESDAPQOQSSPRGKRSRAPPADKFAFAEAQKLPDPPCWTSGLPFPVP----- 1058
Db 1433 TESSMTQFKSKEG-----TELLEITWKETVPETPDHVSSEGEVDVFTLSSHDK 1484
QY 1059 ---PREVIKASHADPSAFSPCHPLPLGLHDTA-----RPVLP----- 1098
Db 1485 TTRWSSEITSSNLENP---VHKQP-KPVLPPESSGEGAEQASQETILSRATEVAL 1540
QY 1099 -----PPTISNPPPLISSAKHPSVLERO-----ICAISSQGMV-----QLHVPY 1137
Db 1541 GKETOQSPTLST--SSILSSSVSNVLEERPLITLIGISQDSNSTIESWVEITPQTKVF 1599
QY 1138 SEHAKAPVGVMTWGLPLPMDPKKLAPFSGVKQB-----QISPRQAGPPESLGVPTAQ 1190
Db 1600 SSSAPI-----IEGSEVENKNKIFNMVTDLPQRPDPTDLS-PLDM 1642
QY 1191 EASVLRGATGALSVPGGSIKTIKIPSTRVPSDSALITYRGSITHGTPADVLYKGTITRIIGED 1250
Db 1643 SKIMITNHHI-----YIPATIAPLDSK-----PSPDARPTTVWNSNSTSEWVSDK 1688
QY 1251 SPRLDRGREDLSLPKHVYIEGKKGHVLSYEGGMSVTQCKEDGRSSSGPPHE---TAAP 1307
Db 1689 S-----FEGRK-----KKENEDBEGAVNAHQGEVRAAT 1717
QY 1308 KR-----TYDMMEGRVGAISASIEGLMGRAIPPERHSPHLKEQHHRGSIQGIPIR 1361
Db 1718 ERSDHLLLPLESSNVDASSDLATWEGFILETTPTESE-----KEMANSTPVFREITGV 1772
QY 1362 SYVEAQ-----EDVLRREAKLLKREGTPPPPPSRDLTEAVKQALGPLKLKAHEGLV 1415
Db 1773 ANVEAQPFHSSSHPRVQELITLSGNPPSL-----FTDLGSGDASTGMELITA----- 1822
QY 1416 ATVKEAGRSIHIPRE-ELRHTEPLPLAPRLKEGS-----ITQGTPLKYD-TGASTT 1466
Db 1823 -----SLFTLDLESETKVKELFTSPSPSVSEISSFEPTGLTPSTVLDEIAGVMSQ 1874
QY 1467 GSKKHDRVSLIGSPKRTPPVHPVLDVADARALERACYEESLKSRPGTASSGSGSIARGA 1526
Db 1875 TSQKTLISEISKP-----TSQSGVRDLTYGF 1901
QY 1527 PVIVELCKPROSLTYEDHGAPFAG---HLPRGSPVTMRBPT-----PRLOEGSL 1575
Db 1902 PM-----GDFSGDFSEYPTVSYPTMKBEETVGMGSDDERVDTQTS 1943
QY 1576 SSKASQDKLSTPREIAKSPHSTVPEHHPHPISPYHLLRGVSG---VDLYRSHI--- 1628
Db 1944 SSIPTSDNIYVPVD--SKGPDSTVAS---TTAFPEWEEVMSAARGSGEQLASRVSSGVP 1998
QY 1629 -PLAFD-----PTSIP---RGIPLDAAAAYLPRHLAPN 1658
Db 1999 LPLAVDIFSGTESPYDFDEFEVAAVTEANERPVLPTAAGNTVDLTENGIEVNTMS 2058
QY 1659 PTYPHLYPPYLIRGVPDTAALE---NRQTIINDVITSQMHNTATAMAQRAQMLRGLSP 1715
Db 2059 LDFPQTEPESKLVGSKPEVNLKQBEIGRETVTKERAAQKQTFESLHSSFAPEQITILETQSL 2118
QY 1716 RESSLALNYAAGPRGIDLSOVPHLPVLVPTTP-----GTPATANDRLAYLPTA 1764
Db 2119 IETFFQIS-----DYSMLTTLKYITNKVEEEMGSIAMHSTPGPGIKOLESYTH 2169
QY 1765 POPFSSRHSSPLSPGGPHTLTKPTTTSSSERDRDRDREREKSLITLTTTVEHA 1824
Db 2170 PEAPCKSHSFS-----ATALVTESGAAR-----SVLMDSSSQEEE 2204
QY 1825 PI---WRPG---TFQSSGSSSGGGGSSSRP-----ASHSHA 1857
Db 2205 SIKLFQKGVKLTNKESSNADISFSLGSGGALPPLPTTSVNLTKQIISTLYAETSHMES 2264

QY 1858 HQHSISPRTQDALQORRPSVLHNTGMKIITAV-----EPSKPTVLRSTSTSSVRPA 1910
Db 2265 LGTSLGDKMED--HERMEDVSSNEVRMLISKIGSIQDSSTEALDITLTSHTGTTEPT--T 2320
QY 1911 ATFFPATPCPLGGTLDGVYPTLMEPVLLPKEAP-----RVARPERPRADTGHAFKAPAR 1966
Db 2321 STLP-----FVKLMDLERSPKQDPGGKRPKTHRPQTMGSLINENSSAS 2366
QY 1967 SGLPSPASPS-----KGSEPRPLVPVPSGHA-----TIARTP 1998
Db 2367 EAREGATSPATFLPQTYSEVMTKHFAPSEOSPLFNVSNGEGSEVDTLDLVYTSQTQ 2426
QY 1999 AKN-----LAPH---HASPPPPAPPASADPHREKTKOSKPSQOELEBLRSLGHVGSYSP 2050
Db 2427 ASSQGSNKLASHGLEKHPKSVKTEAGATDV--SPTASAMF-----LHSEYKSSLYPT 2478
QY 2051 EGVFVPSVPSPLTHDKGLPKHLEELDKGHLGE-LRPKQPGVKVLGGEAAHLPHLRPL 2109
Db 2479 STLSTPEPKSPSEIGIEDGLQ-----DNTQFEGSTLKPSR-----RKT 2516
QY 2110 PES-----QPSSSPLLOTAPGVKHQRVWTLAQHISEVITQDYTRHHPOQ 2154
Db 2517 TESIIDLKEDSKDLGLTITITESAIVKSLPELTSKNI-----IIDIDHTK----- 2562
QY 2155 LSAPLPAPLYSF-PGASCPVLDLRRPPSDLYLPPDPHGAPARSPHSE--GGKRSPEPNK 2211
Db 2563 -----PVYEYIPGIQ---TDL-----DPEIKLESHGSSSESLQVQKVEGAVTILSPTE 2607
QY 2212 TSVLGGGEDGIBFVSPPEGMTEPGHSRSVYVYLLYRDGEOTE----- 2253
Db 2608 ESFEGSGDALL-----AGYTOAIYNES-VTP---NDGQKQAEIISFPATGIPVSSTETE 2657
QY 2254 -----PSRMGSKSPGNTSQP-----PAFFSKLITESAMVKKKQKINK 2293
Db 2658 LHFTFPTASTLHTPSPKLTASP-EIDKPNIEAISLDDIFESSTLSDQAI-ADQSEVIST 2715
QY 2294 LNTHNNEPEYINSQPCTEIFNNPAITGTLTYRSQAVQEHASTNM-----GLEAIR 2347
Db 2716 LGHLEKTOEYEBEKYGGFSQFQEPFSGVGEVLTDPATVSGTSLYIAQTLTELPNVVR 2775
QY 2348 KALMGKYDOWESPPPLSANAFNPLNASASLPAA-WPITAADGRS-----DHLTSPGGGG 2401
Db 2776 PSDSTHYT--EATPEVSSLA---ELSPQIPSSPFVYVNDGVSKFPPEVPHTSIAQP--- 2825
QY 2402 KAKVSGRPSRKA-----KSPAGLASGDRPPSVS-----SVHS 2435
Db 2826 ---VSTVTSQKSIESPKEVHANIEETIKLGGNVHRTPEPMSGRDPALDVSDESKHK 2882
QY 2436 EGCNRRRTLTNRVWEDRPSSA-----GSTPPFPYPLNRLQAGVMAWASPPPLPAGSGP 2490
Db 2883 LLEELETSPTKETSQDFPNKAKDHIPGET-----VGLAGIRTTESEPVITADDME 2934
QY 2491 LAG-----PHHA 2497
Db 2935 LGATQOQPHSA 2945

RESULT 100
NFH_MOUSE
ID NFH_MOUSE STANDARD; PRT; 1087 AA.
AC F19246; Q61959;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neurofilament triplet H protein (200 kDa neurofilament protein)
DE (Neurofilament heavy polypeptide) (NF-H).
GN NEFH OR NFH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Theria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=89121513; PubMed=3220257;
 RA Julien J.-P., Cote F., Beaudet L., Sidky M., Flavell D., Grosveld F.,
 RA Mushynski W.;
 RT "Sequence and structure of the mouse gene coding for the largest
 RT neurofilament subunit.";
 RL Gene 68:307-314 (1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RP MEDLINE=8908138; PubMed=3145094;
 RA Shneidman P.S., Carden M.J., Lees J.P., Lazzarini R.A.;
 RT "The structure of the largest murine neurofilament protein (NF-H) as
 RT revealed by cDNA and genomic sequences.";
 RL Brain Res. 464:217-231 (1988).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP STRAIN=Swiss Webster; TISSUE=Brain;
 RA Carden M.J.;
 RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: Neurofilaments usually contain three intermediate
 CC filament proteins: L, M, and H which are involved in the
 CC maintenance of neuronal caliber. NF-H has an important function in
 CC mature axons that is not subverted by the two smaller NF proteins.
 CC -1- PPM: There are a number of repeats of the tripeptide K-S-P. NFH is
 CC phosphorylated on a number of the serines in this motif. It is
 CC thought that phosphorylation of NFH results in the formation of
 CC intermediate cross bridges that are important in the maintenance
 CC of axonal caliber.
 CC -1- PPM: Phosphorylation seems to play a major role in the functioning
 CC of the larger neurofilament polypeptides (NF-M and NF-H), the
 CC levels of phosphorylation being altered developmentally and
 CC coincident with a change in the neurofilament function.
 CC -1- SIMILARITY: Belongs to the intermediate filament family.
 CC -1- CAUTION: Ref.2 sequence differs from that shown in positions 534
 CC to 716 and is shorter due to frameshifts.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M24496; AAA39813.1; JOINED.
 DR EMBL; M23349; AAA39813.1; JOINED.
 DR EMBL; M24494; AAA39813.1; JOINED.
 DR EMBL; M24495; AAA39813.1; JOINED.
 DR EMBL; M35131; AAA39809.1; JOINED.
 DR EMBL; Z31012; CAA83229.1; -;
 DR PIR; J03068; QPMNSH.
 DR MGD; MGI:97309; Nefh.
 DR InterPro; IPR001664; IF.
 DR Pfam; PF00038; filament; 1.
 DR PROSITE; PS00226; IF; 1.
 KW Intermediate filament; Coiled coil; Neurone; Phosphorylation; Repeat.
 FT DOMAIN 1 97
 FT HEAD.
 FT DOMAIN 98 408
 FT ROD.
 FT DOMAIN 409 1087
 FT TAIL.
 FT DOMAIN 436 517
 FT GLU-RICH (ACIDIC).
 FT DOMAIN 519 886
 FT 50 X 6 AA TANDEM REPEATS OF K-S-P-A-E-A.
 FT DOMAIN 887 1087
 FT GLU/LYS-RICH.
 FT COIL 1A.
 FT DOMAIN 98 129
 FT COIL 1B.
 FT DOMAIN 130 141
 FT LINKER 1.
 FT DOMAIN 142 239
 FT COIL 1B.
 FT DOMAIN 240 261
 FT LINKER 12.
 FT DOMAIN 262 283
 FT COIL 2A.
 FT DOMAIN 284 287
 FT LINKER 2.
 FT DOMAIN 288 408
 FT COIL 2B.
 FT DOMAIN 133 133
 FT K -> OA (IN REF. 2 AND 3).
 FT DOMAIN 199 199
 FT A -> AR (IN REF. 2 AND 3).
 FT CONFLICT 199 199
 FT S -> T (IN REF. 2 AND 3).
 FT CONFLICT 281 281
 FT L -> G (IN REF. 2 AND 3).
 FT CONFLICT 492 492
 FT P -> PREAKSP (IN REF. 3).
 FT CONFLICT 551 551

FT CONFLICT 689 712 MISSING (IN REF. 3).
 FT CONFLICT 714 714 G -> A (IN REF. 3).
 FT CONFLICT 814 814 V -> M (IN REF. 2 AND 3).
 FT CONFLICT 843 843 T -> N (IN REF. 2 AND 3).
 SQ SEQUENCE 1087 AA; 116612 MW; 57BAC76A38ED1CB9 CRC64;
 Query Match 2.3%; Score 309; DB 1; Length 1087;
 Best Local Similarity 21.5%; Pred. No. 0.002;
 Matches 245; Conservative 164; Mismatches 491; Indels 238; Gaps 55;
 Qy 290 ARKQWKQ-----KFCQRYDOL--MEALEKKVRIENNPRRAKESKVREYVEKQFPEIR 341
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 91 ARSEKQQLALNDPAGYIDKVRQLEAHNRSLSEGAALRQOKGRAANGELYE----- 143
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 342 KQRELOERMOSRVQGRSGLSMSAAR-----SEH---EVSEIIDGLSEQ-----ENLRKQ 388
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 144 --REVRE-----MRGAVLRGARGQLRLEQEHLLDIAHVQRQLDEARQREAEAA 194
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 389 MRQLAVIPPMYDADQRIKFINNGLMADPMVKYKORQVNMMSQEKETTFREKFMQHP 448
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 195 ARALA---FAQEAEAAARV-----ELQKKAQAL---QEECGYLR--HHQ 230
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 449 KNFGLIASFLERKTVAECVLYYLTKNVYKSLVRSYRR-RGKSQOQOQOQOQOQOQOQO 507
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 231 EEVGELLGQIQGCGAAQAQ---AARDALUKDVTSAIRAIQAQLEGHAVOSSLSQSEW 287
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 508 QQQPMPSRSEKDEKEKEAEKEEKEPEVENDKEDLLKEKTDITDTSGEDNDEKAVASK 567
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 288 FRVPLDLRSEAAKNTDMSAOE-----EITEYRQLQARTT-----ELEALKS- 332
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 568 GRKTANSQGRKGRH-TRSMANEANSBEATPQOSAEIASMELNESSRWTE-EEMETAKK 625
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 333 ---TKESLQRSELEDRHQADIASYQDAI-QQLDSELNRTKWEAAQALREVQDILLNVKM 388
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 626 GLLHEGRNWSIARWVSKTVSQCNFY--PNYKQRQLDEI--LQOH-KLMEKERNAR 680
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 389 AL---DIEIAAYRLLEGE---ECRIGFGSPFSLTEGLPKIPISITHIKVYSEMIKV 442
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 681 RKKKK---APAAASAEAPPPVVEDEMEASGVSGNEEMVEAEALH-ASGNEVPGRGE 735
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 443 EKSEKTVIVGQTEETRVTEGVTEEDKEAQGEAEAGEEKEEELAAATSPPAEE 502
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 736 CSGPATVNNSDTESIPSPHTEAAKDTGQNGPKPATLGADGPPPGPTTPRRTSRAPTE 795
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 503 AASPEKETKRVKEAKSP--GEAKSPGE--AKSP-----AEAKSPG-----EAKSPGE 547
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 796 PTPASEATGATPPAPPSAPSAPPV-VPRKEEETAAAPVVEGEQKPPAAELAVD 854
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 548 AKSPGEAK-SPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 600
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 855 TGKAEPEPVKSECTEAEAGEGPAKGDAAEAEATAEGALKAEKKEGSGRATTAKSSGAPQD 914
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 601 --EAKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPK 657
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 915 SDSATCSADEVDEAGGDKNRLSPRLTPTGDPANASQKPLDLK---OLKQRAA 971
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 658 SPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEP 703
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 972 AIPPIQTVKHEPPREDAAATKRAPAPPAPPONLQPSDAPQOQSSPRGKS-----RSP- 1026
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 704 AKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 762
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 1027 --APPADKEAPAAEAQKLPD---PPCWTSLGPPFPVPPREVIKASPHAPDPSAFSAPPG 1081
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 763 EVKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPAEPKSPA 822
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 1082 HPLPLGLHDTARPVLPRLP-----PTISNPPPLISS-----AKHPSVLERQIGAISQGM 1129
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 823 KAKEL---DVKSPEAQTVQEEATVPTDIRPEQVKSPAEPKSPAEPKSPAEPKSPAEPKSPA 879
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 1130 SVQLHVPYSEHAKAPVGPVTM-----GLPLPM-----DPKKLAPFSGVKQQLSPRG 1176
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
 Qy 880 KEEVKSVPKBEVKAKPEPKKVEEETLPTPKTEAKESKDEAPKEAPKPKVEEKEKTPTE 939
 Db |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

